

276
1.9.70

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, Social Science Research Council

Associate Editor: E. J. SHOEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Center for Advanced Study
in the Behavioral Sciences

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
New York University

MORTON DEUTSCH
New York University

I. E. FARBER
State University of Iowa

LEON FESTINGER
Center for Advanced Study
in the Behavioral Sciences

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

GARDNER LINDZEY
Center for Advanced Study
in the Behavioral Sciences

DAVID C. MCCLELLAND
Wesleyan University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University



PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6. D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 3440 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is \$16.00; foreign, \$16.50. Single issues are \$3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Social Science Research Council
230 Park Avenue
New York 17, N. Y.

N.B.: CHANGE OF EDITORIAL ADDRESS. Beginning September 1, 1956, manuscripts should be sent to the Editor at the following address: Dr. M. Brewster Smith, Graduate Department of Psychology, New York University, Washington Square, New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

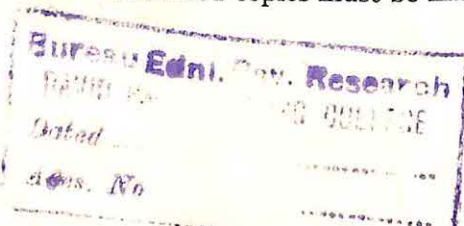
Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland



DISTURBANCES AND SILENCES IN THE PATIENT'S SPEECH IN PSYCHOTHERAPY¹

GEORGE F. MAHL

Yale University

THIS paper presents exploratory research on the measurement of two expressive properties of the patient's speech in psychotherapy and on their use as indices of anxiety. First, general considerations that formed the background of this work will be stated briefly.

Linguistic measures of patient anxiety during therapeutic interviews should be useful in several areas of psychological research. (a) If such measures are correlated with physiological indices, certain psychosomatic relations in human behavior can be more readily investigated. The need to pursue such problems arose during the course of one clinical study (22) and a series of experimental studies (16, 17, 18, 20) of anxiety and HCl secretion in dogs, monkeys, and humans. This need was the original, though now receding, impetus for the present work. (b) Much can be learned about the "psycholinguistics" of anxiety that could be of value to students of language and of cognitive processes. (c) Because anxiety is a nodal variable in psychotherapy, anxiety measures should facilitate the objective investigation of psychotherapy. For example, descriptions of changes during the course of psychotherapy or comparisons of the effect of various therapeutic techniques and procedures will be enhanced if they effectively portray the changes or differences in the patient's anxiety. A review of the literature shows that only a few objective studies of therapy have tried to measure anxiety.

¹ This report deals with one aspect of current research on "The Patient's Language as Expressive Behavior." Most of the work presented here was supported by the Foundations' Fund for Research in Psychiatry. Some of it was also supported by the USPHS, Grant M-1052. The writer wishes to express his appreciation to Richard Karpe, M.D., and John Higgins, M.D., for many important contributions to this program, and to Robert Abelson, Ph.D., for advice on statistical techniques.

A more extended discussion of this exploratory work, including a manual of instructions for training people to score speech disturbances and a circuit diagram of the timing apparatus, is available (19). The high-fidelity facilities used to obtain the tape recordings have been described elsewhere (21).

The preceding interests require the measurement of anxiety as it occurs *immediately* in the therapeutic session. This process is different from determining characteristic anxiety reactions, defenses, etc., in the patient's general behavioral repertory such as one attempts for diagnostic purposes. For the present goals, one wants to assess the *current* behavioral state of the patient.

In the therapeutic interchange, the intensity of the patient's anxiety may fluctuate markedly during the course of a single interview. Continuous change must also be anticipated in most of the variables with which one would wish to correlate patient anxiety, such as physiological measures, various categories of therapist behavior, patient-content categories, or other linguistic measures. As a result, procedures giving only single measures of anxiety for large segments of interviews or for entire sessions taken as the units might obscure more precise covariations that inevitably would be of interest. It seems that the most generally useful procedure would involve the measurement of some relatively continuous aspect of the patient's behavior. The data could then be summated or averaged for whatever temporal or behavioral units were required for specific problems.

Sound recordings supply one with a continuous stream of behavior, the language of the participants, from which linguistic measures of anxiety meeting the preceding requirements can potentially be derived. The basic working hypothesis underlying the present use of recordings for *this* purpose has been that the most valid linguistic measures of anxiety will be those based on the behavioral or "expressive" aspects of the speech rather than those based on manifest verbal content analysis. A full discussion of the reasoning back of this assumption is beyond the scope of the present paper, but some of it appears in the discussion. This working assumption, of course, is one that must be re-examined during the research in progress.

Empirically, two of the many behavioral

attributes of speech in the interview that are useful to the therapist in assessing anxiety in the patient are (a) disturbances in speech called "jumbled," "confused," or "flustered" speech, and (b) hesitations and longer silences by the patient when he is free and motivated to talk. Theoretically, silence (and perhaps speech disturbance) may be regarded as a defense motivated by anxiety evoked by ideational events or by the nature of the interpersonal relation. Speech disturbances and short hesitations may also be conceived as predominantly indirect linguistic consequences of anxiety that do not have the instrumental function of reducing anxiety. This notion is based on the assumption that one effect of anxiety, regardless of its source, is to disrupt *all* complicated ongoing behavior, irrespective of its behavioral relation to the source of the anxiety. Here, speech is merely an excellent instance of such complex behavior susceptible to the disruptive effect of concurrent anxiety.

This exploratory research was intended to devise reliable methods of measuring these two attributes of speech from sound recordings of interviews, to study the general discriminatory properties of the measures, and to explore a procedure for validating these measures as indices of patient anxiety.

THE MEASURES

Speech-Disturbance Measure

Upon analyzing instances of obvious speech confusion in interviews, it is found that they consist of a variety of separate identifiable forms of disturbances occurring at a rapid rate. These same disturbances occur at lower rates at times when the patient's speech is not clinically experienced as confused. A quantitative measure of this variability can be obtained by identifying the various disturbances in a verbatim transcript and then computing the following ratio for any given language sample:

Speech-Disturbance Ratio =

$$\frac{N \text{ Speech Disturbances}}{N \text{ "Words" Spoken by Patient}}$$

In this paper Speech-Disturbance Ratios for successive two-minute segments of therapeutic interviews and for certain content and interaction phases of interviews are considered.

In practice the speech-disturbance categories are marked in the transcript while listening to the recording *and* reading the transcript. The "word" count used to date has included the number of completed words, the number of incomplete words, the number of "sounds" caused by stuttering and incoherent sounds, and the number of "ah's." The value placed in the denominator could probably just as well be some other measure of the amount of talking done by the patient—such as the "number of seconds of patient talk." The requirement is only to have a disturbance measure that is independent of the verbal output by the patient.

The speech-disturbance categories presently used are listed below and briefly defined. These were developed empirically by studying recordings and typescripts, and noting disturbances in individual words and in the word-word progression that were *sensed* as superfluous or distorting to the communication of the content. Judging from experience in scoring thirty-odd interviews for twelve different patients and speech samples of 40 undergraduates, this set seems to be generally adequate.

Disturbance Categories

1. *"Ah."* Wherever the definite "ah" sound (as distinguished from "er," "um," etc.) occurs, it is scored.
2. *Sentence Correction.* A correction in the form or content of the expression while the word-word progression occurs. To be scored, these changes *must be sensed by the listener* as an interruption in the word-to-word sequence.
3. *Sentence Incompletion.* An expression is interrupted, clearly left incomplete, and the communication proceeds without correction.
4. *Repetition.* The serial superfluous repetition of one or more words—usually of one or two words.
5. *Stutter.*
6. *Intruding Incoherent Sound.* A sound which is absolutely incoherent as a word to the listener. It merely intrudes without itself altering the form of the expression and cannot be clearly conceived of as a stutter, omission, or a tongue-slip (though some may be such in reality).
7. *Tongue-slip.* This category includes neologisms, the transposition of words from their correct serial position, and the substitution of an unintended for an intended word.
8. *Omission.* Parts of words, or rarely entire words, may be omitted. Contractions are exempted. Most omissions are of terminal syllables of words.

Real examples of these disturbance categories are contained in the following interview

excerpt. In the transcript an asterisk (*) is placed at the point of the disturbance. To the left are the categories in the corresponding line of the excerpt. If there is more than one disturbance in a line, the category names are in sequence with the asterisks. Since the actual scoring is done only when listening as well as reading, some of it may not be completely clear from this written material alone. Poor grammar, inaudibility of the recording, and interruptions by the therapist are not grounds for scoring the categories.

Illustrative Scored Excerpt

Abbreviations

1. In transcript: P = patient; T = therapist
2. In left margin: SC = sentence correction
R = repetition
St = stutter
O = omission
T-S = tongue-slip
Inc = incompleteness
IS = incoherent sound

T-S, SC

Ah

2 St

Ah, R

R

T-S, O, SC

Ah

Ah

Ah

P: My impression of my relation to D-(son) have always been that the reason that I don't* . . . *didn't seem to feel the love for him that I felt for J-(daughter) was that during the first sixteen months of his life I was away. I didn't grow up with him. If there was any jealousy of D- it was in relation to his in-laws. Now that is very possible. Although it's something which I also suppressed. Ah* . . . and the reason I say it's possible 'cause it sort of well le.* le.* leaves a . . . ah* . . . a* sort of memory. When I (?) I . . .

T: You're jealous of his in-laws?

P: Of his in-laws, yeah. Because he was brought up with them till . . . until* I came home. He was born in their hou** . . . *hospital and came to their house, and my wife lived with her parents.

T: Ho . . . how do you mean you're jealous of his in-laws?

P: Well . . . ah* . . . when I first came home, and for the first year or so, or more than a year . . . ah* . . . D- was more prone to turn to his grandfather and grandmother than he was to me.

T: Mmmhnn.

P: And although I understood it, there was a certain amount of . . . ah* . . . well not bitterness, I wasn't bitter about it, but a certain

SC

Inc

SC, St

Ah

Ah

R

IS, O (?), IS, St

O

SC, Ah, SC

amount of . . . *a sort of resentment, a mild type of resentment.

T: You make this sound so . . . ah-

P: Well-

T: So diluted.

P: Well it was diluted. I mean it wasn't something which I felt keenly enough to be angry at his grandparents, let's say, or with D-himself. I mean I realized that he had grown up with them.

T: Mmmhnn.

P: And therefore, it was more natural for him to . . . *until he became . . . *completely o.* overcame his . . . ah* . . . strangeness to me, and it took quite a long time. Ah* . . . then there was . . . *there was a certain amount of resentment. It wasn't directed against the parents or dec* . . . *was . . . den* . . . ac.* actually directed against a circumstance which kept me away. And it was a resentment which re* . . . *in a certain . . . ah* . . . *to a certain extent reflected itself also in the (clears throat) feeling I had toward people who had remained behind and had made money.

T: Mmm. Your in-laws make money?

P: Yes. My father-in-law made a lot of money from the war.

The Patient-Silence Quotient

A Patient-Silence Quotient is the silence measure described here, although other means of assessing silence have been explored in this research. For any given segment of an interview this is defined as follows:

Patient-Silence Quotient =

$$\frac{N \text{ Seconds of Silence}}{N \text{ Seconds Available to Patient to Talk}}$$

Silence Quotients for successive two-minute segments in the interview and for certain content and interaction phases in the interview are considered later in the paper.

The basic apparatus for obtaining the relevant activity data consists of a tape recorder and a pair of hand switches each connected with the clutch of a Springfield timer. The observer's task is to listen to the recording and to press one switch whenever and for as long as the patient talks and to press the other switch in comparable fashion for the therapist's talking. Each clock runs accumulatively whenever its hand switch is closed. At the end of the interview segment

involved, the total number of seconds each participant was observed to be talking is read from the clocks. The values for computing the Silence Quotient for the segment are then derived as follows:

- $a = N$ seconds in the segment
 $b = N$ seconds talk by therapist
 $c = N$ seconds talk by patient
 N seconds of silence = $a - (b + c)$
 N seconds available to patient = $a - b$

When successive two-minute Silence Quotients are desired, apparatus with the following operating features is used. The observer momentarily closes a foot switch which starts the tape recorder, a telechron motor geared for $\frac{1}{2}$ r.p.m., and which closes relay switches in series with the clutches of the accumulating clocks. He then operates the hand switches as described above. A microswitch is automatically opened when the telechron motor completes one revolution in two minutes. This instantly releases the relays to stop the tape recorder and inactivate the clock clutches. The clock readings are recorded, the clocks are reset to zero, and this sequence repeated for the next two minutes, etc. By using this apparatus the raw data for the two-minute Silence Quotients of a 45-60 minute interview can be obtained in about one and a half hours by an experienced operator.

To obtain the Silence Quotients for the larger and variable time segments reported below, there are some minor changes in procedure and use of the apparatus which will not be described here.

Reliability of the Measures

Three people have learned to score speech disturbances. The interobserver reliability has been determined by correlating the total N of disturbances scored by independent judges in unselected samples of transcript pages ($N = 28-65$) for several different patients ($N = 3-5$). The average product-moment intercorrelation was .94.

One observer has learned to obtain Silence Quotients with the apparatus and procedure described above. Test-retest runs at two to four week intervals were made for five interviews. Three interviews were for one patient, and one each were from two other patients. The average product-moment correlation of

the test-retest two-minute Silence Quotients was .96.

GENERAL PROPERTIES OF THE MEASURES

Thirty-odd interviews from twelve patients have been scored for speech disturbances and measured for Silence Quotients. Most of these interviews have come from two patients undergoing psychotherapy. The remainder of this paper pertains to the material from these two patients. This section of the paper treats only the successive two-minute measures.

Tables 1 and 2 contain descriptive statistics of the two-minute measures for Mrs. Y and Mr. Z. Speech disturbances and silence are prominent modes of behavior for these patients from the purely quantitative standpoint. On the average, Mrs. Y produces one speech disturbance per 23 spoken "words" and Mr. Z one speech disturbance per 16 "words."

TABLE 1
SUMMARY OF TWO-MINUTE SPEECH-DISTURBANCE RATIOS FOR MRS. Y AND MR. Z

Mrs. Y				Mr. Z			
Interview No.	N 2' Ratios	Mean	SD	Interview No.	N 2' Ratios	Mean	SD
1	24	.037	.019	2	30	.057	.021
2	25	.041	.014	15	30	.068	.023
5	30	.038	.015	17	31	.060	.040
10	33	.052	.018	20	29	.062	.026
13	28	.042	.015	26	32	.064	.034
17	28	.049	.026	67	30	.055	.029
20	26	.036	.013	83	30	.075	.024
26	29	.048	.024	88	30	.050	.027
Mean = .043 .018				Mean = .061 .028			

TABLE 2
SUMMARY OF TWO-MINUTE SILENCE QUOTIENTS FOR MRS. Y AND MR. Z

Mrs. Y				Mr. Z			
Interview No.	N 2' Ratios	Mean	SD	Interview No.	N 2' Ratios	Mean	SD
5	30	.260	.253	2	30	.370	.133
10	29	.291	.122	5	30	.467	.160
13	28	.383	.137	15	30	.455	.071
17	28	.445	.220	17	30	.579	.196
20	26	.304	.224	20	29	.554	.099
26	29	.308	.173	26	30	.587	.137
				67	30	.567	.163
				83	30	.516	.094
				88	30	.570	.134
Mean = .332 .188				Mean = .518 .132			

On the average, Mrs. Y is silent 33 per cent of the time she could talk and Mr. Z, 52 per cent of the time he could talk. These are not atypical patients in these respects.

Tables 1 and 2 suggest that these measures differentiate therapeutic interviews with respect to mean level and variability. Figures 1 and 2 contain graphs of the successive two-minute Speech-Disturbance Ratios and Patient-Silence Quotients for Mrs. Y's 17th session. These graphs illustrate an impression one gets that the measures tend to portray trends or phases within many interviews, as well as differentiating between interviews for a given patient. Confirmation of these points would be compatible with the belief that the scores were behaviorally meaningful in general, and not just artifacts of the laboratory.

Bartlett's test and a simple analysis of variance were used to test for the between-hour differences in variability and means. The results presented in Table 3 show that there are significant between-hour differences in both variability and means for both measures. Since the variances and the sample N 's meet the conditions for which no appreciable error is made in interpreting the F test (10), it can be concluded that the differences in means are true differences and not due to the within-interview heterogeneity of variance.

The Wald-Wolfowitz run test was used to evaluate the presence of phases in the two-minute graphs. This test is described by Moses (24). Briefly, in the present application, the median of the two-minute measures for each hour is determined. A "run" is defined as a series of one or more scores falling on the same side of the median. The number of runs for an interview is determined and the difference between the observed and the theoretically expected number is measured in standard error units. Under the null hypothesis, in a given sample of interviews, the sum of the deviate scores resulting from the run test would be zero. The standard error of the sampling distribution of such sums would be \sqrt{N} . The significance of the departure of the obtained sum from zero is then evaluated with the normal probability table.

The hypothesis tested here was that in each patient's sample of interviews there would be on the whole fewer runs, for both measures, than if the successive two-minute measures were random uncorrelated scores. The hypothe-

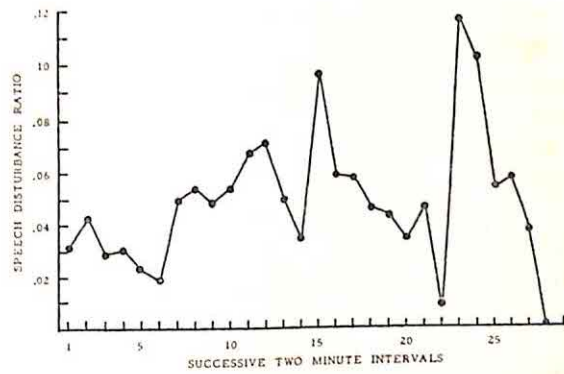


FIG. 1. SPEECH-DISTURBANCE RATIOS IN SUCCESSIVE TWO-MINUTE INTERVALS OF MRS. Y'S 17TH SESSION

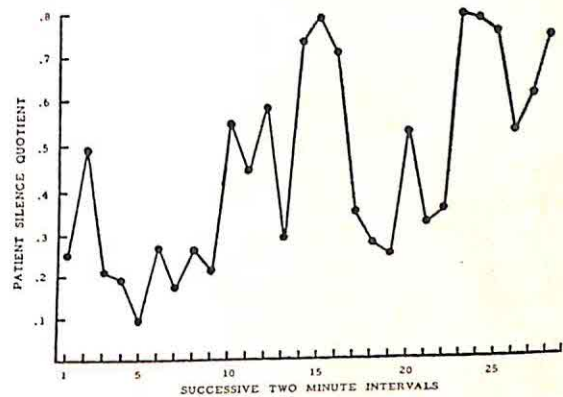


FIG. 2. PATIENT-SILENCE QUOTIENTS IN SUCCESSIVE TWO-MINUTE INTERVALS OF MRS. Y'S 17TH SESSION

TABLE 3
TESTS OF BETWEEN-INTERVIEW DIFFERENCES IN TWO-MINUTE MEASURES

Measure	Patient	Bartlett's Test		Analysis of Variance	
		χ^2	p	F	p
Speech disturbance ratio	Mrs. Y	24.39	<.001	2.7	<.02
	Mr. Z	16.14	<.05	2.19	<.05*
Silence quotient	Mrs. Y	11.18	<.05	5.55	<.001
	Mr. Z	37.22	<.001	9.88	<.001

* The violation of the assumption of homogeneous variance might cause this to be spuriously "significant." The true value might be as high as .07 (see 10).

sis of too few runs was posed because of general knowledge about the interaction between the therapist (the writer) and these patients. He allows the patient to become anxious as a result of his association trends and the interaction, but he also permits the patient to resort to defensive maneuvers for varying periods of time depending on the circumstances. The

therapist also at times consciously intervenes to reduce the patient's anxiety. At times he is insensitive to the patient's anxiety, in which case the patient usually succeeds in successfully maneuvering as long as the therapist follows the lead of the patient. At times the therapist changes the topic himself because he too becomes anxious. Of course anxiety is not always reduced either by the patient or the therapist, so every hour would not be expected to show clear phasic changes in a purported anxiety measure.

One of Mr. Z's interviews, the 83rd, was omitted from this test. This was because over

a year before any of this work started, the writer has evaluated it and used it for demonstration purposes because of its illustration of sustained conflict with many oscillations in anxiety level. Thus, it was already known that it would be inappropriate to include this hour in the present test of too few runs.

The results of the Wald-Wolfowitz tests presented in Tables 4 and 5 confirmed the prediction of few runs for both measures in both patients. An obvious refinement of this test would be to predict independently which interviews would show more and which less than the number of runs expected by chance.

A PROPOSED VALIDATION PROCEDURE

Preliminary Findings

The measures of speech disturbance and silence were selected for study because there were empirical and theoretical reasons for believing they would be valid linguistic indices of concurrent anxiety in the speaker. The preceding findings show that these two speech attributes discriminate something in the patient's behavior. The questions now are: is it variability in anxiety that is being discriminated and how can this be determined?

During therapeutic sessions and while studying recordings it often appears that interviews are divisible into natural segments or phases, each of which could be assigned to a single theme of content or interaction, and that the patient becomes anxious and conflictful in some, but becomes less anxious in others. In some of the latter, it can often be observed that the patient himself has changed the topic or started a new line of interaction with the therapist in such a way that his behavior can be interpreted as a relatively sustained and successful *defensive maneuver*.

During live interviews or in listening to recordings, judgments of such changes in anxiety as have been mentioned are based in large part upon changes in expressive speech attributes. The assumption underlying the validation procedure and test to be described is that given adequate context for interpretation, it would be possible to judge such phases in typescripts only (not recordings). If this were done validly, and if the Speech-Disturbance Ratio and Silence Quotient are valid indices of patient anxiety, then these two measures should be greater in the anxious or

TABLE 4
RESULTS OF WALD-WOLFOWITZ RUN TEST APPLIED TO SUCCESSIVE 2-MINUTE SPEECH-DISTURBANCE RATIOS

Mrs. Y		Mr. Z	
N Obtained- N Expected Runs*	z	N Obtained- N Expected Runs*	z
-7.5	-2.95	-3.0	-1.11
-1.5	-.57	-4.0	-1.48
-4.0	-1.48	-.5	-.18
-.6	-.27	-6.0	-2.22
-3.0	-1.2	-4.0	-1.48
-3.0	-1.2	+3.5	+1.32
+.5	+.2	+2.0	+.72
+1.5	+.53		
$\Sigma = -6.94$ $p < .01$ (one tail)		$\Sigma = -4.43$ $p < .05$ (one tail)	

* For individual interviews.

TABLE 5
RESULTS OF WALD-WOLFOWITZ RUN TEST APPLIED TO SUCCESSIVE 2-MINUTE SILENCE QUOTIENTS

Mrs. Y		Mr. Z	
N Obtained- N Expected Runs*	z	N Obtained- N Expected Runs*	z
-9.0	-3.59	-9.0	-3.33
-5.0	-1.92	-7.0	-2.59
-3.0	-1.11	-7.0	-2.59
-4.5	-1.70	-1.0	-.37
-1.0	-.38	-8.0	-2.96
-.5	-.19	-9.0	-3.33
		-4.5	-1.70
		-7.0	-2.59
$\Sigma = -8.89$ $p < .001$ (one tail)		$\Sigma = -19.46$ $p < .001$ (one tail)	

* For individual interviews.

conflictful phases than in the low anxious or defensive phases.

A test of these predictions has been made for Mrs. Y, using interviews 5, 10, 13, 17, 20, 26 as the test hours. Before the test was made, *ad hoc* clinical evaluations of a similar nature with interviews from Mr. Z and from other patients appeared to give promising results. In addition, almost the same experimental-clinical procedure described below had been followed with still different interviews of Mr. Z from his first 26 sessions. The results were essentially the same as for Mrs. Y, and are presented elsewhere (19). They are not described here because the possibility of contaminated phase judgments was not as well controlled as in the test with Mrs. Y's interviews.

With Mrs. Y, the first major problem was to prevent contaminated criterion judgments. The writer had been the therapist, but the test interviews occurred slightly over two and one-half years before this validation test was made. The writer had not previously replayed the recordings or studied transcripts of these interviews. After practicing with other interviews, a secretary edited verbatim transcripts and prepared clean scripts that did not contain any speech disturbances, any annotations concerning pauses, or any explicit reference to silence by the participants. These scripts were no longer exactly verbatim, of course. With these precautions, it was felt that no serious contamination of the phase judgments occurred.

The judgment of the proper anxiety category for the phases of content or interaction requires intimate knowledge of the context. One of the main reasons for this is that the judgment cannot be made on the basis of manifest content alone. The context includes general knowledge of the patient and of the dynamic setting of a given therapeutic session. The development of the context is a major problem. A good deal was known about the patient since the judge had been exposed to over 100 therapeutic interviews with her. A more immediate context of the test hours was obtained by listening to the recordings for the first 29 interviews, *excepting the test hours*. Notes made of each of the nontest hours and the complete edited typescripts were then reviewed in sequence. Before the final judging of phases in any given test hour, a rough clinical description of the therapeutic situation at the time was written.

After the interviews were divided into phases

TABLE 6
SPEECH-DISTURBANCE RATIOS AND SILENCE QUOTIENTS
IN PHASES OF INTERVIEWS. MRS. Y

Measure	Statistic	Phase		Significance (one tail test)
		Low Anxious or Defen- sive	Anxious or Con- flictful	
Speech-disturb- ance ratio	N	15	15	
	Mean	.038	.049	$t = 2.68$ $p < .01$
	SD	.013	.009	
Silence quotient	N	15	15	
	Mean	.316	.427	$t = 3.51$ $p < .001$
	σ^2	.0032	.0122	$F = 3.94$ $p < .01$
	SD	.054	.107	

and categorized as to anxiety type, the phases were marked off in the original verbatim typescripts. A second person then determined the Silence Quotients and the Speech-Disturbance Ratios for the phases. Just as the phase judge had no prior knowledge of the objective measures, the scorer of the objective measures had no knowledge as to the identity or meaning of the phases.

A total of nineteen defensive and/or low anxiety and nineteen anxious or conflictful phases were judged in the five test hours. For any given interview, however, there had not always been judged an equal number of each type of phase. To control for the confounding of "hour effects" demonstrated earlier, equal numbers of the phase types were randomly selected for each interview. This procedure produced a sample of fifteen phases in each group. Table 6 contains the pertinent statistical data. Both objective measures were significantly higher in the anxious or conflictful phases. The mean Speech-Disturbance Ratio increases by 29 per cent, while the mean Silence Quotient increases by 35 per cent in the anxious over the other phases. It will be noted that the Silence Quotients were also significantly more variable in the anxious phases. Nevertheless, the p value of the mean difference is such that one can safely conclude that there is a significant difference in means as well as in variability.

Illustrative Interview

The material for Mrs. Y's 13th interview illustrates the kind of interpretive phase judgments involved and the way in which the linguistic measures vary with the

phases. Before proceeding to the interview itself, some preliminary information concerning the patient is presented.

Mrs. Y, a college graduate, was in her late thirties at the onset of treatment. She had been married for eleven years. She had two daughters, aged $7\frac{1}{2}$ and 9. A third pregnancy had terminated with a miscarriage some seven months earlier. The patient sought treatment ostensibly for relief from recurring duodenal ulcers and because of a general feeling of unhappiness and dissatisfaction in her relations with her husband and children. This was consciously related to frustration of strong dependency needs based on a lifelong history of repeated rejection by both of her parents as well as well as other significant people. As therapy progressed, a strong sex-fear conflict with a great deal of Oedipal flavoring seemed equally, if not more, important in determining her present troubles. This conflict was not only specific to genital behavior but was generalized to many modes of feminine behavior. For example, even though she moved in a fashionable circle and was not restricted in income, she was afraid to buy and wear the kinds of attractive clothing worn by her peers. She always chose "practical" and "adequate" clothes. While she was aware of "missing the boat" in her sexual relations with her husband, the sexual conflict and the general fearful avoidance of femininity and womanhood was almost completely unconscious in the early stage of therapy.

At the time of the 13th session the patient had been in treatment for a month. During this period she had manifested and felt a positive attachment to the therapist. She was conscious of the dependency aspect of this attachment but not of an inferred unconscious sexual aspect. Consciously she had felt frightened at the dependent attachment, being afraid that the therapist would reject her as her father had done irrevocably and traumatically when she was ten, and as she alleged a previous therapist had done.

There were some unusual aspects of her treatment arrangement that are also pertinent to the evaluation of the 13th session. She knew the therapist was doing research on both recorded therapy interviews and on the psychophysiology of peptic ulcer patients. She also knew that her therapy was included in these studies and that the therapist had accepted her for treatment partially for these reasons. She was aware of the investment of time and energy in each case and the importance of each patient for the research program. In fact, she had felt this as an attraction to enter into treatment with this particular therapist.

Just before the 13th session the patient had been informed by her obstetrician that she was pregnant. The interview is centered on this topic. Becoming pregnant raised many psychological problems for her, but two aspects are particularly prominent in this particular interview. First, she tells the therapist that she is definitely pregnant and, second, she and the therapist explore an irrational fear she has of telling her mother that she is pregnant.

Prior to judging phases in the hour, it was known that both of these aspects were anxiety-laden matters for her. She had suspected some two weeks earlier that she might be pregnant and at that time awoke during the nights before her 8th and 9th sessions preoccupied

in part with how the therapist would react to this news. She anticipated that if she were pregnant her therapy would be interrupted and that this would let the therapist down in his research program. Her thoughts at 3:30 A.M. on the day of the 9th session were:

"Should I tell Dr. Mahl?" She anticipated that he would think "That's just like a woman. There goes my research program. She is gumming up the works." She also said to herself, "Your research program depends in part on me."

Also in the 9th interview she stated that she was afraid to tell her mother she might be pregnant. After the 13th interview her mother came to visit for several days as had been previously arranged. The day her mother came, the patient became "ill" and had to go to bed. Only after this did she tell her mother she was pregnant. (The mother reacted positively.) These observations show that the irrational fear of telling her mother which is explored in the 13th interview was real and intense.

The 13th session. This interview is now summarized by phases. For each phase, the content or interaction is first summarized, then the category into which the phase was placed and the reasons at the time of the judgment for doing so are stated. The Speech-Disturbance Ratios and Silence Quotients for the phases are presented in Fig. 3.

Phase 1

After spending 30 seconds in idle chitchat about the weather, Mrs. Y abruptly comments: "I got off to a bum start this morning. I'm mad! At least I got mad at my family before breakfast even." She proceeds to state that she woke up with a hollow feeling in her stomach and ate a cracker. She casually interposes the news that since last time she has discovered she is definitely pregnant because she "... had an A.Z. test, whatever that is." She immediately goes on to describe in detail her "bum start." Her two daughters had provoked her, one by getting toothpaste on a clean dress and the other by putting on a clean dress instead of one she had worn before. The patient had become irritated and her breakfast is "sitting right there in a lump." She closes this account by saying that these are more unnecessary things she has to cope with.

Interpretation of the phase. It is inferred that the patient is afraid of the therapist's reaction to the definite knowledge that she is pregnant. The nature of this fear is the same as she described earlier in the 9th session. This anticipatory fear motivates her to inform the therapist of the A.Z. test results as casually as possible and to induce a sympathetic attitude in him by portraying herself as the "victim" of her daughters. The content of the phase is instrumental in doing these things. By being receptive to the account, the therapist "aids" the patient in doing these things. For these reasons, this is judged to be a relatively *successful defensive maneuver* in which the anxiety of the patient is minimal. Both linguistic measures were at their lowest level in this phase.

Phase 2

The patient spontaneously changes the subject saying straightforwardly now that she had the A.Z.

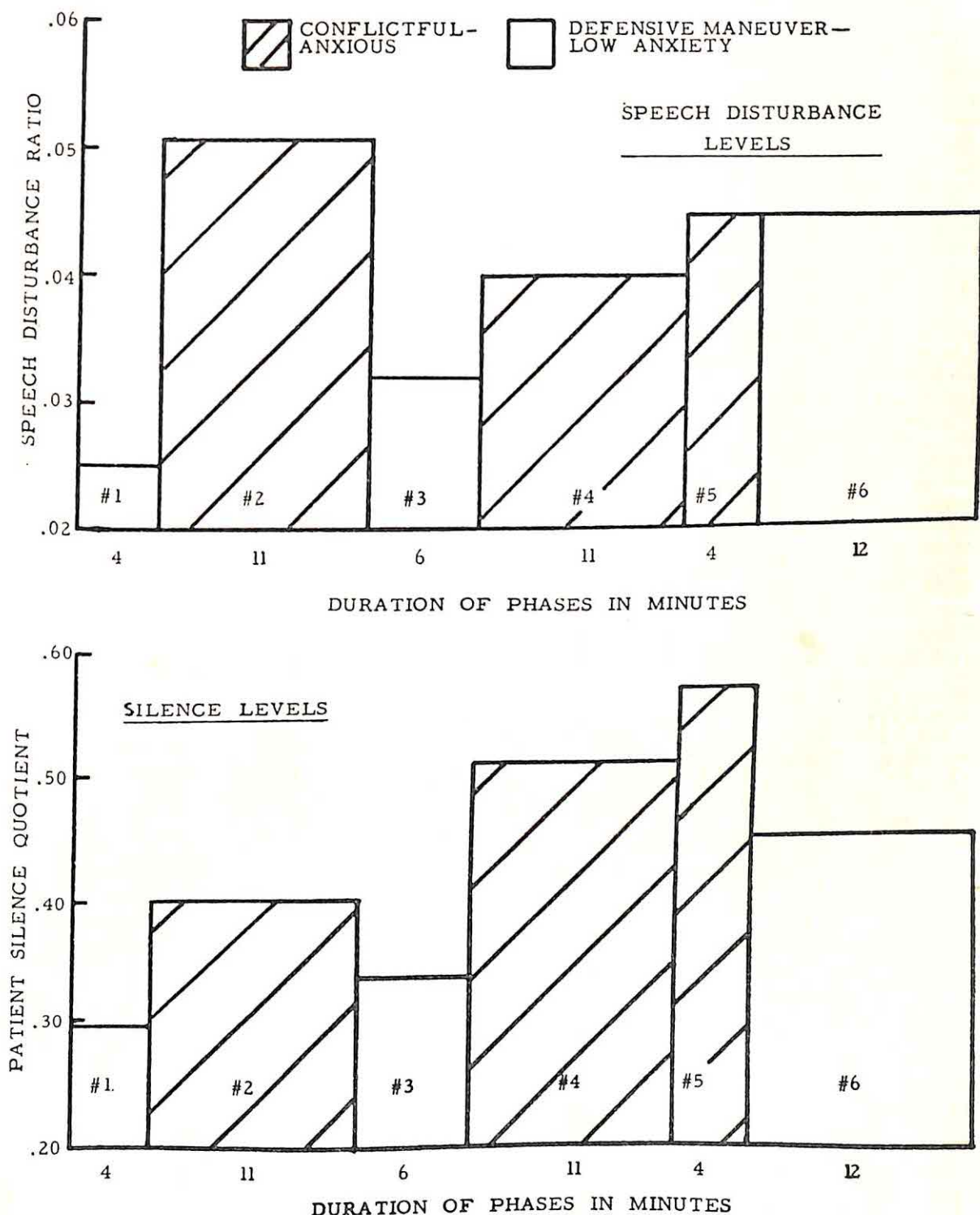


FIG. 3. SPEECH-DISTURBANCE RATIOS AND PATIENT-SILENCE QUOTIENTS IN THE SUCCESSIVE PHASES OF MRS. Y'S 13TH SESSION

test and it was positive. Then she says, "Now I'm trying to figure out how I'm going to get through having my mother with me for a week without telling her I'm pregnant. 'Cause I don't want to tell her." She elaborates very briefly saying she knows her mother will be

upset and will think the patient shouldn't have any more children, especially so soon after the miscarriage.

The therapist returns to her news of being pregnant by asking the patient how she felt when she heard the A.Z. was positive. She replies that it struck her as

amusing because she conceived immediately after she and her husband decided to try again. She adds that she'll be glad to have another child. The therapist asks how this jibes with earlier statements by her to the contrary. With many contradictions, she says that she is glad because her husband wants another child, but that it's not her favorite idea and she wouldn't do it if she didn't feel she should. The patient comments briefly on how she is resigning from several community activities in order to keep her activities at a minimum and prevent another miscarriage.

The therapist asks then for further elaboration of her fear of her mother's reaction. The patient says again she is afraid to tell her mother and that she was always afraid to do so with her previous pregnancies. The therapist remarks: "That's striking, isn't it." The patient acknowledges this and states that while she did, and does, feel afraid, her mother has always expressed surprise when the patient tells her of this fear. Even so, the patient continues, she knows her mother really does disapprove and that she will be critical of her. But then she expresses the belief that there must be something more to this feeling than she has said. The therapist agrees. The patient says she doesn't know what the basis for this feeling could be and would like to know. The therapist states that by thinking about her relationship with her mother now she might start to find out.

Interpretation of the phase. This is regarded as a phase in which the patient starts to talk about the anxiety-evoking news that she is definitely pregnant. This is the material she was avoiding in part, and preparing for in part, in the first phase. Three things happen that increase her anxiety: (a) she tells the therapist directly that she is pregnant, (b) under some pressure from the therapist, her surface "happiness" at being pregnant is rendered ineffectual as a defense against the anxiety and guilt elicited by her negative feelings, and (c) the defensive aspect of blaming her mother for the anticipatory fear at telling her of being pregnant is partially exposed.

As can be seen in Fig. 3, both the speech-disturbance and silence measures increase during this phase.

Phase 3

Responding to the remark of the therapist made at the end of Phase 2, the patient now introduces the theme that her mother has never seemed maternal or supportive, especially when something happens that increases demands upon her. The patient elaborates this with rather detailed descriptions of two instances in which her mother had responded to illness in the family by seemingly becoming upset, excited, impatient, and anxious to escape the situations as soon as possible.

Interpretation of the phase. This is interpreted as a rather extended defensive maneuver consisting of a well-organized rationalization of her fear of the mother's reaction. Since this is a reinstitution of a defense that had just previously been partially exposed, the anxiety that this reinstitution of the defense was necessary here contributed to the judgment of Phase 2 as conflictful or

anxious.) It will be seen in Fig. 3 that the behavioral-linguistic measures both decrease during this phase from their level in Phase 2.

Phase 4

The therapist asks how her mother had reacted to the patient's previous pregnancies in such a way as to account for her present fear. The patient describes her mother as having been interested in the pregnancies, having brought gifts, etc. She repeats how her mother had expressed surprise at the patient's fear of her reaction. The therapist comments that this suggests there hasn't been any basis in her mother's behavior to justify her fear and that maybe the source of the fear lies within herself. The patient asks if the therapist means a feeling between herself and her mother or between herself and the rest of the world. After the therapist replies that he doesn't know the answer to this and that it must come from her, Mrs. Y says: "I'd like to know. I really would like to figure out what it is. I mean whether I'm transferring to my mother some feeling that I have, whether it's a feeling of guilt about having them or what, I don't know." She elaborates this by saying the guilt might be because she doesn't think she is a good mother to her children. The therapist asks if this is an attempt to intellectualize and says this particular feeling could not have been present when she was afraid to tell her mother the first time she became pregnant. Mrs. Y agrees and wonders why she should feel guilty about being pregnant. Shortly she reports feeling blocked in thinking further about it. (It had been previously determined that reports of blocking by the patient are not themselves reliable clues of silence. Therefore such ambiguous references were left in the script.)

Interpretation of the phase. This is a conflictful or anxious phase. The patient should be highly conflictful because the therapist is continually confronting her with the inadequacy of her previous accounts of why she fears her mother's reaction. The patient momentarily gives up the defensive rationalizing, is left blocked in this mode of defensive behavior, and feels repressed near the end of the phase.

It is seen in Fig. 3 that the Silence Quotient rises markedly in this phase. Observation of the recording and the unedited typescript after the phase judgment had been made showed that most of the silence occurred at the end of the phase. Thus, it instrumented the repression or suppression the patient alluded to. The Speech-Disturbance Ratio also rises in this phase.

Phase 5

When the patient says she feels blocked at the end of Phase 4, the therapist states that this may often happen in her therapy and that while he could offer his thoughts about her fear of her mother, it really wouldn't help. Mrs. Y replies that she realizes that he probably doesn't want to plant ideas in her mind and that "I've gotta dig for it." She goes on to say that this fear of telling her mother is funny, because she doesn't mind telling other people about being pregnant. She feels guilty about telling her mother and she shouldn't

have to feel this way. But this is true of other things too, for she hasn't told her mother very much about herself for years.

Interpretation of the phase. This was judged to be another highly conflictful phase. The therapist was pressing the patient to continue without her defensive maneuvers, and the patient made open acknowledgments that the source of the fear was within herself. In so doing, she was more directly giving up her prepotent defenses.

Both measures rose in this phase, although the rise over Phase 4 had not been specifically predicted. That they should do so makes sense *ad hoc* because the therapist was actively prodding the patient when she was already anxious and conflictful.

Phase 6

The patient states that previous attempts to confide in her mother were "flops." When the therapist asks what happened, Mrs. Y responds by recalling in some detail an incident in her teens when she was in boarding school. She had told her mother that she and a boy were corresponding with each other. Shortly afterwards the boy unaccountably stopped writing and the patient was sure her mother had interfered, maybe because she did not approve of men. With collaboration from the therapist, the patient describes other instances of her mother interfering with her dating. In general she depicts her mother as having disapproved of all such adolescent heterosexual activities.

Interpretation of the phase. This was judged to be a defensive phase in which the patient's anxiety should be decreased from the previous anxiety phase. The content was regarded as a *displaced reference* to the basic unconscious fear of giving expression to her feminine sexual desires and needs.

In Fig. 3, the Silence Quotient decreases in this phase, but the speech disturbance does not decrease. This is commented upon in the discussion.

DISCUSSION

References to silence, blocking, and confused speech by the patient are numerous in clinical records and reports. But this is not the case in the scientific literature on objective research with recorded interviews.

To the writer's knowledge there has been no systematic analysis of the disturbed speech phenomena considered in this paper in therapeutic interviews. Some investigators have measured verbal activity or word output in interviews of various types and in relation to numerous variables (1, 2, 3, 4, 5, 6, 8, 11, 12, 13, 14, 15, 26, 27, 30). Of these, only Ruesch and Prestwood (30) deal directly with anxiety in interviews. Discrete silences *per se* were made the explicit object of study by Tindall and Robinson (34). They conducted a descrip-

tive study of silences in nondirective interviews, covering such factors as who initiated the silences, their apparent meaning, and their effect on the progress of the interview. Gillespie (9) included the frequency of discrete silences in a resistance measure. Silence was measured implicitly, of course, in most of the verbal activity studies. But the activity measures have usually been such that they are not simply the obverse of the degree of silence.

Chapple (2) devised a method for obtaining activity measures in interviews that required an observer to operate hand switches when the participants spoke. These switches activated magnetic markers that recorded on polygraph paper. A variety of temporal measures could then be made from such recordings, including instances and durations of simultaneous silence by both participants. Verzeano and Finesinger (35) later developed a speech analyzer that eliminated the need of the observer under certain conditions and that yielded automatically much of the data available with Chapple's method. The method and apparatus used here to obtain the Silence Quotient are merely further refinements on Chapple's idea for a special purpose.

Thus, there is general awareness that silence and disturbances are significant aspects of the patient's speech. Yet, as the summary review above shows, there has been relatively little investigation of these attributes in general and almost none of their specific use as indices of patient anxiety. The work described in this paper shows that the Speech-Disturbance Ratio and the Silence Quotient are reliable and discriminating measures, and suggests that there is cause for further studies of their validity as anxiety indices.²

There were several reasons for using motivational phase judgments as validation criteria. Physiological measures usually seem "safer," but they involve their own methodological problems, often serious ones, and the psycho-

² There is also an unexpected side-product of the analysis of confusion in speech into the discrete disturbance categories. Individuals appear to differ in the degree to which they use the various categories and there are some lawful relations between the categories. Some of these findings were presented at the Eastern Psychological Association meeting, 1955, in a paper "The use of 'ah' in spontaneous speech" and are being prepared for publication now.

physiology of emotions is not so well established as to provide one with ready-made measures guaranteed to be generally applicable indices of various emotions. Clinical ratings of interviews are also candidates for validation criteria. Even when shown to be reliable, they threaten validation studies because they cover large units of time within which there may be considerable behavioral variation. The motivational phases are behaviorally as meaningful as physiological measures; they focus on smaller units of behavior than ratings of interviews; and they prevent errors arising from being "content bound."

In the specific case of linguistic measures of anxiety, the defensive maneuver and conflictful phases seem especially useful as validation criteria. They are significant and frequent occurrences in psychotherapy. Sullivan, for example, presents a very useful discussion of them (33). This study might be regarded as an exploration, suggesting a procedure for establishing adequate context for their judgment and for preventing contamination. A more adequate design, of course, would include the determination of the reliability of the phase judgments. It was a calculated risk to proceed here with this gap in design, for negative findings would have been ambiguous concerning the anxiety-measuring property of the Speech-Disturbance Ratio and Silence Quotients.

In the introduction it was stated that the basic working assumption of this research was that the most valid linguistic measures of anxiety will be those based on "expressive" aspects of speech rather than on manifest content analysis. The basis for this assumption must be briefly considered. First, however, what is meant here by "expressive" aspects of speech may be clarified. In a speech sample there are verbal (words) and vocal elements. The vocal elements consist of *conventional vocal symbols* and *vocal clues or signs* of emotional change that may be perceived by a therapist but were not intended as symbols by the patient. The former are illustrated by culturally determined intonation patterns that serve a function analogous to punctuation marks in writing. The latter are illustrated by the speech disturbance categories described here. They might be compared to variations in writing pressure or letter formation that are unknown

to or unintended by the writer but could be detected by a measuring technique. The term "expressive" aspects of speech is used here to refer to only this second group of vocal attributes.

Manifest content analysis of interviews yields quantitative descriptions of what the patient says. As generally practiced, however, small content units in written material are categorized in an uninterpretive way in the sense that the instrumental function of the content and the expressive properties accompanying the content *at the time* it is uttered are ignored. (Probably the most extreme instances of such work is that in which individual statements of patients have been typed on separate cards and these *written* and *isolated* utterances used to measure affective states.) If one uses such manifest content scores obtained by uninterpretive means as indices of emotional states, he assumes that there is more often than not some sort of uniform ("between" and "within" patients) and usually direct reflection of an emotional state in the words uttered. A simple example would be measuring currently present fear in the patient by counting the number of sentences he utters containing fear words. At times this may be a valid procedure. But there are also many instances in which a person will not utter fear sentences when he is frightened and may utter them when he is not frightened. Illustrative occasions in therapy are those in which a patient, frightened by a therapist, may verbally attack him in content, though in an uncertain voice, and those in which a very dependent patient feels calm and protected and sounds comfortable, while engrossing the therapist with accounts of fearful events.

If the disparity between the manifest content, on the one hand, and the motivational context and the expressive aspects, on the other hand, were relatively infrequent in psychotherapy, one could proceed safely to use content analysis in measuring emotions. But this type of disparity *seems* too frequent to disregard. It is certainly great enough in both everyday life and in psychotherapy to have received special consideration. Sapir (31) was referring in part to this distinction in his contrast between language as an expressive as well as a symbolic-referential system. George H.

Mead (23) made this distinction in his lucid contrast of a "conversation of gestures" with a "conversation of significant symbols." In the clinical literature, Reich's (29) classical distinction between "what" and "how" a patient speaks, and Sullivan's (33) contrast between "vocal" and "verbal" behavior both call attention to the same problem.

Probably all aspects of speech are subject to individual differences in learning and emotional usage. Thus, such expressive measures as Speech-Disturbance Ratios and Silence Quotients may be subject to the interpretive approach just as manifest content is. This seems most likely with silence, which at times may be a result of anxiety but at other times may be a form of stubbornness or hostility, of thoughtfulness, seductiveness, etc. To the extent that this is so, the routine use of speech disturbances and silence measures as indices of anxiety is subject to the same limitations as content measures. This research work is simply based on the starting belief that (a) the expressive measures are less influenced by the learning process in socialization and are thus less variable in their meaning and less subject to the errors of the uninterpretive approach than manifest content, and (b) inferences about the behavioral state of the patient based on uniform treatment of them will be right significantly more often than wrong. Only further work will determine if this belief is correct.

In the meantime it seems likely at least that the Speech-Disturbance Ratio and the Silence Quotient could be profitably used in conjunction with simultaneous manifest content measures like the Distress-Relief Quotient of Dollard and Mowrer (7), the "feeling" scores of the nondirective therapists (28, 32 e.g.), and the more refined content measures of Murray (25). These expressive measures might reduce the error in inferring behavioral states from the verbal content even though they do not turn out to be superior or adequate by themselves in this respect. This potentiality may be illustrated with Phases 3 and 6 in Mrs. Y's 13th session. In Phase 3 the patient accounts for her fear of telling her mother she is pregnant by referring directly to her mother's failure to gratify people's real needs for support and indirectly to her own fear of dependency deprivation. In Phase 6, the patient refers to her

mother's disapproval of her dating boys in her adolescence. This was regarded as a displaced reference to the patient's fear of reproval for the feminine sexual manifestations involved in pregnancy. On the basis of content alone, it would be very difficult to state in which phase this patient was more successfully defending against anxiety-evoking thoughts. The expressive measures suggest that she was decidedly more anxious in Phase 6 than in Phase 3 in that the displacement in Phase 6 was a less successful defense than the rationalization of Phase 3. In the course of her therapy sexual fears and conflicts have been the source of much greater resistance than have dependency fears and conflicts.

The effect on these measures of other variables than motivational states must still be assessed. This seems particularly true with respect to the speech-disturbance measure and the effect of verbal intelligence, education, social class, and sex. One gets the impression that these variables do not seriously limit the value of the Speech-Disturbance Ratios in making comparisons "within" patients, but systematic evidence is lacking. Also it is likely that a way to combine the two measures into a single one would be useful. Since silence seems best regarded as a defense, the Silence Quotient may increase due to mounting anxiety and yet reduce the Speech-Disturbance Ratio. Figure 3 suggests that this type of interaction may have occurred in the conflictful phases of Mrs. Y's 13th session. This raises a difficult problem of weighting the two measures. Scores obtained by simple *z*-score combinations have not seemed as useful as the individual measures taken separately in some preliminary attempts along these lines.

SUMMARY

Linguistic measures of the patient's anxiety in therapeutic interviews can be useful in several areas of research. A working assumption is that the most valid measures will be based on the expressive aspects of speech rather than on manifest content measures. On empirical and theoretical grounds, speech disturbance and silence seem to be expressive attributes that are useful as anxiety indices. Reliable methods for measuring these two attributes have been described. Data from two cases show that the

two measures discriminate "something" between-sessions and within-sessions for a given patient. Motivational phase judgments have been proposed as useful criteria for validating the Speech-Disturbance Ratio and Silence Quotient as anxiety measures. Detailed knowledge of the dynamic context of the interview is necessary to make such judgments of criteria since manifest content is of limited value in this regard. A procedure for acquiring the context without contaminating these judgments has been described. Preliminary results of the proposed validation procedure indicate that the present approach is fruitful. The research and ideas presented here are exploratory and can only be regarded as providing leads that might be of interest to others working in this area so costly of research time.

REFERENCES

1. CARNES, E. F., & ROBINSON, F. P. The role of client talk in the counseling interview. *Educ. psychol. Measmt.*, 1948, **8**, 635-644.
2. CHAPPLE, E. D. Measuring human relations: an introduction to the study of interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3-147.
3. CHAPPLE, E. D. "Personality" differences as described by invariant properties of individuals in interaction. *Proc. nat. Acad. Sci., Wash.*, 1940, **26**, 10-16.
4. CHAPPLE, E. D., & HARDING, C. F., III. Simultaneous measures of human relations and emotional activity. *Proc. nat. Acad. Sci., Wash.*, 1940, **26**, 319-326.
5. CHAPPLE, E. D., & LINDEMANN, E. Clinical implications of measurements of interaction rates in psychiatric interviews. *Appl. Anthropol.*, 1942, **1**, 1-11.
6. COBB, S. Presidential address. *Trans. Amer. neurol. Ass.*, 1949, 1-8.
7. DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3-32.
8. FINESINGER, J. E., LINDEMANN, E., BRAZIER, MARY A. B., & CHAPPLE, E. D. The effect of anoxia as measured by the electroencephalogram and the interaction chronogram on psychoneurotic patients. *Amer. J. Psychiat.*, 1946-47, **103**, 738-748.
9. GILLESPIE, J. F., JR. Verbal signs of resistance in client-centered therapy. In *Group report of a program of research in psychotherapy*. Pennsylvania State Coll., 1953. Chap. IX.
10. JONES, L. V. Statistical theory and research design. *Annu. Rev. Psychol.*, 1955, **6**, 405-430.
11. LASSWELL, H. D. Verbal references and physiological changes during the psychoanalytic interview: a preliminary communication. *Psychoanal. Rev.*, 1935, **22**, 10-24.
12. LASSWELL, H. D. Certain prognostic changes during trial (psychoanalytic) interviews. *Psychoanal. Rev.*, 1936, **23**, 241-247.
13. LINDEMANN, E., & FINESINGER, J. E. The effect of adrenalin and mecholyl in states of anxiety in psychoneurotic patients. *Amer. J. Psychiat.*, 1938-39, **95**, 353-370.
14. LINDEMANN, E., & FINESINGER, J. E. The subjective response of psychoneurotic patients to adrenalin and mecholyl. *Psychosom. Med.*, 1940, **2**, 231-248.
15. LORENZ, M., & COBB, S. Language behavior in manic patients. *Arch. Neurol. Psychiat.*, 1952, **67**, 763-770.
16. MAHL, G. F. Effect of chronic fear on the gastric secretion of HCl in dogs. *Psychosom. Med.*, 1949, **11**, 30-44.
17. MAHL, G. F. Anxiety, HCl secretion, and peptic ulcer etiology. *Psychosom. Med.*, 1950, **12**, 158-169.
18. MAHL, G. F. Relationship between acute and chronic fear and the gastric acidity and blood sugar levels in *Macaca mulatta* monkeys. *Psychosom. Med.*, 1952, **14**, 183-210.
19. MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. Unpublished progress report, January, 1955.
20. MAHL, G. F., & BRODY, E. B. Chronic anxiety symptomatology, experimental stress, and HCl secretion. *Arch. Neurol. Psychiat.*, 1954, **71**, 314-325.
21. MAHL, G. F., DOLLARD, J., & REDLICH, F. C. Facilities for the sound recording and observation of interviews. *Science*, 1954, **120**, 235-239.
22. MAHL, G. F., & KARPE, R. Emotions and HCl secretion during psychoanalytic hours. *Psychosom. Med.*, 1953, **15**, 312-327.
23. MEAD, G. H. *Mind, self, and society*. Chicago: Univer. of Chicago Press, 1934.
24. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
25. MURRAY, E. A case study in a behavioral analysis of psychotherapy. *J. abnorm. soc. Psychol.*, 1954, **49**, 305-310.
26. PAGE, H. A. An assessment of the predictive value of certain language measures in psychotherapeutic counseling. In *Group report of a program of research in psychotherapy*. Pennsylvania State Coll., 1953. Chap. VII.
27. PORTER, E. H. The development and evaluation of a measure of counseling interview procedures. Part II. The evaluation. *Educ. psychol. Measmt.*, 1943, **3**, 215-238.
28. RAIMY, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153-163.
29. REICH, W. On character analysis. In R. Fliess (Ed.), *The psychoanalytic reader*. New York: International Universities Press, 1948.
30. RUESCH, J., & PRESTWOOD, A. R. Anxiety. *Arch. Neurol. Psychiat.*, 1949, **62**, 527-550.
31. SAPIR, E. Language. *Encyclopedia of the social sciences*. Vol. 9. New York: Macmillan, 1933. Pp. 155-169.

32. SNYDER, W. U. An investigation of the nature of non-directive psychotherapy. *J. gen. Psychol.*, 1945, **33**, 193-223.
33. SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954.
34. TINDALL, R. H., & ROBINSON, F. P. The use of silence as a technique in counseling. *J. clin. Psychol.*, 1947, **3**, 136-141.
35. VERZEANO, M., & FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 1949, **110**, 45-46.

Received August 4, 1955.

ONE DETERMINANT OF THE CAPACITY TO FREE-ASSOCIATE IN PSYCHOTHERAPY¹

MAURICE K. TEMERLIN

University of Oklahoma

WHEN at the outset of analytically-oriented psychotherapy a client is asked to verbalize everything that comes into his mind, without organizing his thoughts and no matter how trivial, embarrassing, illogical, painful, or confused his thoughts may seem, he then is forced to respond to a situation which for him is almost totally without structure. From his point of view, the instructions to free-associate create an unstructured situation because he is asked to say *everything* that comes into his awareness but he is not told what to say *specifically* or *how* to say it. Moreover, he is not told in what sequence he "should" bring forth the content of his awareness, or what feeling, gesture, tone of voice, or emotion "should" be correlated with his utterances. Finally, he frequently does not know what his responses mean or how the therapist is interpreting them, and often he can perceive no relationship between attempting to free-associate and receiving the help for which he originally asked.

Within this framework it is axiomatic that individuals differ greatly in their capacity to approximate free association at the outset of treatment, with few persons actually being able to associate freely. Too, the instructions to free-associate usually produce a wide variety of defensive reactions on the part of the client. If there are idiosyncratic, internally consistent aspects of personality which are revealed under conditions of reduced external structure (as in the case of projective tests, for example), then individual differences in the capacity to free-associate at the outset of treatment should be correlated with individual differences in behavior in other unstructured situations.

Since the autokinetic situation as it has been used in recent studies (1, 2, 4) tends to be as free of external standards and criteria as the

therapeutic situation described above, the variability of behavior in the early stages of psychotherapy should be correlated with the variability of successive judgments of movement in the autokinetic situation. Specifically, it was hypothesized that clients whose behavior when asked to free-associate is highly variable, flexible, or productive exhibit significantly greater variability when judging the extent of autokinetic movement on successive trials than Ss whose behavior at the outset of psychotherapy is highly inflexible, stereotyped, or rigid.

METHOD

Subjects

Seventeen Ss were selected from a population of 180 neurotic college students who had requested intensive, outpatient, analytically-oriented, personal psychotherapy from the Guidance Service of the University of Oklahoma between 1951 and 1953. Three categories descriptive of behavioral variability, flexibility, or productivity during the first 20 hours of psychotherapy were defined as follows:

Category 1. Extremely Flexible and Productive Behavior. When asked to free-associate, the client throughout the first 20 hours consistently behaves in ways that the therapist considers extremely variable, flexible, or productive. Examples: The client assumes personal responsibility for the conduct of his therapy and works hard to free-associate; he produces material which is meaningful; he interacts emotionally with the therapist and "lives out" emotional experiences during the hour; he demonstrates a willingness to come to terms with his resistances; he gives interpretations a fair hearing, neither rejecting them blindly and dogmatically nor accepting them passively and compliantly.

Category 2. Extremely Rigid and Unproductive Behavior. When asked to free-associate, the client throughout the first 20 hours consistently behaves in ways that the therapist considers extremely inflexible, stereotyped, or rigid. Examples: The client consistently "blocks" and is unable to express himself; he seems almost totally unwilling or unable to express the content of his awareness; he passively and objectively describes his experience in an affectless, uninvolved, or abstract manner; he assumes no responsibility for the conduct of the therapeutic hour; he demonstrates an inability or an unwillingness to come to terms with his resistances; he does not interact with interpretations.

Category 3. Miscellaneous Behavior. Behavior which does not clearly and consistently fall within either of the above categories.

¹ This study was one part of a doctoral dissertation in the Department of Psychology at the University of Oklahoma. Grateful acknowledgement is made for help given by Dr. William B. Lemmon and for advice and apparatus given by Dr. Muzafer Sherif. Mr. Bernard Moskowitz provided statistical consultation. Dr. Alfred F. Glixman and Mr. Robert S. Wilson read the manuscript and made many valuable criticisms and suggestions.

The author, who served as Therapeutic Practicum Supervisor at the Guidance Service, classified each of the 180 members of the population within one of these categories. His judgments were based upon previously obtained records of staff conferences, supervisory or consultant contacts, or practicum course presentations. The therapist of each *S* was then asked to classify his client (or clients) within one of the three categories on the basis of his personal observations of the client's first 20 hours of therapy. Both therapist and experimenter had to agree on each person's classification; and if either therapist or experimenter had any doubt or disagreement whatsoever regarding the client's classification, he was placed in Category 3 and not used as a subject.

This procedure yielded 7 *S*s in Group 1 and 43 *S*s in Group 2. Of the total population of 180, only the 7 persons classified in Group 1 were judged able to approximate relatively free association during the first 20 hours of psychotherapy. The 43 members of Group 2 were then matched as closely as possible with the 7 *S*s of Group 1 in terms of intelligence, sex, socioeconomic status, and the number of hours of psychotherapy experienced at the time the autokinetic data were collected. The two groups then had these characteristics:

Group 1. There were 5 males and 2 females in this group. Ages ranged from 18–30 with a median of 22.0 years. Their IQ's ranged from 117 to 132 with a median of 124. The number of hours of psychotherapy each *S* had experienced at the time the autokinetic data were collected ranged from 21 to 46 with a median of 29.

Group 2. There were 7 males and 3 females in this group. Ages ranged from 18.5–32 with a median of 23.6 years. Their IQ's ranged from 118 to 130 with a median of 125.6. The number of hours of psychotherapy ranged from 22 to 67 with a median of 33.² All *S*s in both groups were of upper-middle socioeconomic status.

Apparatus

The apparatus was the same as that used by Sherif (4). In a dark and soundproof room a pin point of light was exposed through a circular hole one millimeter in diameter. The appearance of the light, the exposure time and the interval between successive exposures were

controlled automatically by a series of relay circuits and were constant for all *S*s on each trial.

Procedure

The *S* was told:

We are trying to perfect a test which will predict how well a person is going to do in psychotherapy. However, your only task will be to tell me how far a point of light moves. I am going to take you into a dark room and show you a point of light. Your job will be to judge the distance the light moves, making your judgments as accurate as possible. I want your eyes to get accustomed to the dark, so I want to blindfold you here and lead you into the dark room.

The statement that the experiment was an attempt to predict therapeutic success was an effort to produce the kind of ego-involved uncertainty frequently experienced by persons beginning psychotherapy.

The *S* was then blindfolded and led by the hand over a long (75 yards) and complex route and seated in the experimental room. This procedure was designed to produce greater spatial disorganization and thus maximize individual differences. It has been demonstrated that individual differences in the variability of judgments of autokinesis increase with experimentally-induced uncertainty (4).

In the darkened room the blindfold was removed, and these instructions were given:

Periodically, I will show you a point of light. After a few seconds the light will start to move. As soon as you see the light move, press this button. A few seconds later the light will go out. As soon as the light goes out, tell me in inches the total distance the light moved, making your judgments as accurate as possible.

Fifty judgments were taken from each *S* in this manner: Three seconds after the *S* had signified that he had perceived movement, the light was turned off automatically. The *S* then voiced his judgment and the light automatically reappeared 60 seconds later.

Treatment of the Data

The average deviation from the median judgment was determined for each distribution of 50 judgments and this score was used as the measure of each *S*'s judgmental variability. The significance of the difference between the autokinetic variability scores of the two groups was determined with the *U* statistic of Mann and Whitney (3).

² This study was conceived and executed within an analytically-oriented framework where the therapeutic process is thought to take from 300 to 500 hours. In these terms, the data of this study were collected at the outset of treatment, and it was felt that none of the *S*s had shown any significant personality changes as a function of psychotherapy.

RESULTS AND DISCUSSION

The variability scores for the 7 Ss of Group 1 ranged from 2.18 to 6.08 with a median of 4.74. The variability scores for the 10 Ss of Group 2 ranged from .37 to 2.44 with a median of .89. Eight of the 10 Ss in Group 2 achieved a lower variability score than that obtained by any S in Group 1.

The Mann-Whitney U test was calculated for $m = 10$ and $n = 7$ as suggested by Moses (3). A U value of 1.15 was obtained which is 4.92 standard deviations from the mean of the U distribution. This value is significant well beyond the .01 level of confidence.

The highly significant difference between the autokinetic variability scores of the two groups lends strong support to the hypothesis of a relationship between the variability of behavior in these two unstructured situations. This appears consistent with the theory of projective tests that there are internally consistent, idiosyncratic aspects of personality which are revealed under conditions of reduced or ambiguous external structure.

These results are considered empirical evidence that those (few) people who possess the ability to render relatively free associations at the outset of analytically-oriented psychotherapy appear to have a greater capacity to experience visual apparent movement in the unstructured autokinetic situation than people whose stream of consciousness flows less freely under the same conditions. It seems probable, in the light of the projective hypothesis, that these results may be interpreted as a function of the tolerance an individual has for experiencing himself. When a neurotic who has requested psychotherapy is asked to free-associate, he is forced by the unstructuredness of the situation to respond essentially to an actual or anticipated experiencing of himself. If he follows the instructions to free-associate, his own attitudes, motives, memories, perceptions, thoughts, and feelings—of which the self is an organization—will sooner or later constitute the content of his awareness and be verbalized. The interpretation in terms of tolerance for experiencing the self is thus consistent with the

common clinical observation that the freer an individual is simply to relax and to "be himself," the more variable and spontaneous is his behavior. On the other hand, individuals with less capacity to "be themselves" appear to exert greater control over their inner experience with a resultant decrease in the variability or spontaneity of their behavior.

Two qualifications require explicit statement. The rigorous method of selection limited the sample to a small number of highly select Ss whose representativeness may not be sufficiently broad to permit wide generalization. Second, it is important that these results not be interpreted in terms of therapeutic outcomes, for the problem of prognosis was considered peripheral to this study. Furthermore, Ss within both groups eventually deviated from the pattern they had demonstrated consistently during the first 20 hours of psychotherapy.

SUMMARY

Those Ss who were unusually variable, flexible, and productive when first asked to free-associate in psychotherapy were compared with a matched group of extremely rigid, inflexible, and stereotyped Ss with respect to the variability of their perceptions of autokinesis. Significant differences were found. It was concluded that the greater an individual's tolerance for experiencing himself, the more variable and spontaneous is his behavior in unstructured situations.

REFERENCES

1. KELMAN, H. C. Effects of success and failure on suggestibility in the autokinetic situation. *J. abnorm. soc. Psychol.*, 1950, **45**, 267-285.
2. KLEIN, G. S., & SCHLESINGER, H. J. Perceptual attitudes toward instability: I. Prediction of apparent movement experiences from Rorschach responses. *J. Pers.*, 1951, **19**, 289-302.
3. MOSES, L. E. Non-parametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122-143.
4. SHERIF, M., & HARVEY, O. J. A study in ego functioning: Elimination of stable anchorages in individual and group situations. *Sociometry*, 1952, **15**, Nos. 3-4.

Received July 5, 1955

A TEST FOR SPECIFIC AND GENERAL BEHAVIORAL EFFECTS OF INFANTILE STIMULATION WITH SHOCK IN THE MOUSE¹

WALTER C. STANLEY

Brown University

AND

JOHN A. MONKMAN

University of Illinois

PROBABLY the fundamental assumption of personality theory in general and psychoanalytic theory specifically is that the helplessness of infancy and childhood breeds personality disorder. For example, White states:

It takes high-powered anxiety to make a neurosis. Acute combat stress or similar traumatic situations can do it in adults. Otherwise only the panics and apprehensions that occur in the helpless years of childhood are sufficiently high-powered (8, p. 258).

A statement such as the foregoing implies a causal relationship between helplessness and severe anxiety. And, in such a context, we may interpret helplessness in at least two ways: (a) as objective helplessness, e.g., the organism is repeatedly exposed to "painful" or "frightening" stimulus situations, but the termination of such aversive stimuli (i.e., reinforcement) is causally² unrelated to any specific response; or (b) as a "sense" or "feeling" of helplessness, e.g., the organism's recognition of its objective helplessness.

It should be quite obvious that this second meaning of helplessness and its presumed relationship to severe anxiety cannot be directly tested in infrahuman animal subjects, simply because such subjects cannot tell us whether they recognize their helplessness. However, the first meaning of helplessness and its assumed relationship to anxiety can be tested with such animal subjects. Accordingly, the present experiment was designed to test the hypothesis that exposure of the infant mouse to arbitrary shock (i.e., shock which the organism could not terminate by some specific response) would lead to greater fearfulness later on than exposure to shock whose offset was made contingent on a specific response.³ The measure of emotionality or fearfulness utilized was the open-field test

(1, 2). As adults, the mice were also given avoidance training trials to assess the theoretically more transferable or generalizable effects of the infantile treatments.

METHOD

Subjects

The Ss were offspring of genetically homogeneous mice (*dba* strain, subline 1) obtained from the Roscoe B. Jackson Laboratory in Bar Harbor, Maine. Eighty-four Ss who received differential infantile treatment were available for later tests. Of these, 36 were tested only in the open field, and 48 were tested both in the open field and in avoidance learning. In addition, data on performance during infantile stimulation with shock were available on nine Ss who did not survive for the later tests.

Apparatus

Infantile shock apparatus. The apparatus used for infantile stimulation consisted of a wooden box, 9¾ in. long, 4 in. wide, and 2¼ in. deep. The shock grid was 5 in. long and was made of ¼-in. bars spaced ¼ in. apart. To minimize S's being trapped in either corner of the shock compartment, a U-shaped piece of Lucite formed the actual walls of this compartment, thus eliminating its two corners. A sliding door, lined with Lucite, was used to bar entrance to the safe side. Shock to the grid was 105 v., with a fixed resistance of 500,000 ohms in series with the grid.

Open field test. This apparatus was like that used by Lindzey (5). It consisted of a circular table, 30 in. in diameter, which was covered with white oil cloth. It was illuminated by a 150-watt bulb suspended 36 in. above its center.

Adult avoidance-training apparatus. This apparatus consisted of a wooden box whose inside dimensions were 30 x 4¾ x 6 in. deep. The entire apparatus was covered with hinged Lucite lids. The grid compartment was 20 in. long; the safe compartment 10 in. long. A guillotine door separated the two compartments. The floor sections of both compartments were hinged to provide, along with appropriate holding relay circuits, for onset of shock 1.5 sec. after S was placed on the grid and for the activation of a Springfield timer which yielded a measure of S's running time to the safe compartment.

Procedure

The over-all procedure is summarized in Table 1. Further details of the design and procedure follow:

Housing. Throughout their lives, Ss lived in conventional two-compartment wooden mouse boxes. A constant artificial light-dark cycle of 14/10 hr. was maintained through automatic control of an overhead

¹ About two-thirds of this experiment was carried out at Brown University by the second author as an Honors Thesis under the direction of the first author who prepared this report.

² To simplify presentation of the problem we omit consideration of the obviously interesting case where fortuitous reinforcement may generate "superstitious" behavior.

³ Cf., the method of Mowrer and Vick (4).

fluorescent fixture. Room temperature approximated 67°, varying from 62° to 72°, and relative humidity varied from 21 to 40%.

Infantile stimulation. The Ss within litters were randomly assigned to the differential infantile treatments with sex balanced insofar as possible. The three types of treatment were: (a) exposure to shock whose offset was contingent on S's moving to the safe end of the apparatus (response-contingent group); (b) exposure to shock equal in duration on each specific trial to the amount of shock received by the matched litter mate in the response-contingent group (arbitrary shock group); and (c) exposure to the shock apparatus with matched trial durations but with shock turned off (the no-shock group).

The Ss received five trials per day for four days. Intertrial duration was 1 min. Thus, each response-contingent S completed its daily session before its mates in the arbitrary and the no-shock groups were run. However, all Ss in a litter were removed from and returned to the nest at the same time to equalize durations of absences from the nest. Timing was done with a stop watch.

Open field test. All Ss tested received a 2-min. exposure once a day for 10 consecutive days. Approximately two-thirds of the Ss were lifted by hand from their living cages and placed on the table top; the remaining one-third were released through specially built doors in the living cages and thus were not handled by E except after each 2-min. test period. Testing order of Ss within litters was balanced from day to day. Defecations were recorded by E.

Adult avoidance training. The Ss were run in trios of litter mates consisting of a response-contingent, an arbitrary shock, and a no-shock mouse. The interval between placement of S in the start box and the onset of shock was 1.5 sec. The duration of intertrial intervals was about 6 min., with all Ss in a trio completing Trial 1 before any received Trial 2, etc. The S in a trio run first was balanced among the three groups. Each S was run to a criterion of five successive avoidances in a single training session. The E also recorded response time, the time required for S to reach the safe compartment, to the nearest .1 sec.

It is important to note that the shock apparatus and procedure used with adult mice were different from those used with infant mice in the response-contingent group. There were two reasons for this, one practical and one theoretical. The practical reason was that the infant apparatus and training were adapted to a small, nonseeing (6), nonjumping, crawling, and essentially hairless animal, while the adult apparatus and training were adapted to a larger, seeing, jumping, swiftly darting, and hair-covered animal. The theoretical reason was that our major hypothesis required escape conditioning operations in infancy, but our check on the effects of infantile escape conditioning, we assumed, would be most sensitive if avoidance conditioning operations were used. Solomon and Wynne (7) have shown in dogs that a phase of escape learning precedes avoidance performance. Accordingly, it was felt that any effect of infantile escape conditioning operations would be most likely to be demonstrable by giving adult Ss avoidance training directly in a simple alleyway apparatus.

RESULTS

The infantile training data for the response-contingent group were analyzed to determine whether any improvement in performance had occurred. None had. For example, the median daily response times, based on each S's median response time for its five trials, were 8.5, 8.8, 10.0, and 9.8 sec. for an *N* of 37. Perhaps the clearest indication of chance changes in performance came from a comparison of each S's response time on its very first trial with its response time on its last trial. If we include the two Ss who received only 15 infantile training trials, the *N* for this comparison is 39. Nineteen Ss performed better and 20 Ss performed more poorly.

The open-field data are summarized in Table 2. It is evident from the table and statistical evaluations that the three types of infantile experience had no differential effect on excretory behavior in the open field, other than what may be attributed to chance.

TABLE 1
OUTLINE OF PROCEDURE

Age in Days	Treatment
7*	Toes clipped for identification
8-11†	Infantile exposure to shock apparatus
30	Weaning
45-54‡	Open-field test of emotionality
59-133, <i>M</i> = 90	One day of avoidance training to criterion

* Day 8 for 6 Ss.

† Days 9-11 for 6 Ss.

‡ Days 42-51 for 6 Ss; Days 35-44 for 3 Ss.

TABLE 2
MEAN DAYS OF DEFECTION IN THE OPEN-FIELD TEST

Total <i>N</i>	Group			$\sigma_{M-diff.}$ §
	Response-Contingent Shock Offset	Arbitrary Shock Offset	No-Shock	
14 trios*	3.1	3.5	3.3	.82, 1.24, .95
24 trios†	4.5	3.7	3.6	.84, .91, .84
30 pairs‡	4.1	3.8	—	.74

* Later tested for avoidance learning (see Table 3).

† Inclusive of the 14 trios.

‡ Inclusive of 24 pairs from the 24 trios.

§ Respectively, for the response-contingent vs. arbitrary shock, the response-contingent vs. no-shock, and the arbitrary shock vs. no-shock comparisons involving litter-matched Ss.

TABLE 3
PERFORMANCE IN ADULT AVOIDANCE TRAINING

Individual Measure	Total N	Group					
		Response-Contingent Shock Offset*		Arbitrary Shock Offset		No-Shock*	
		Mdn.	Q_U and Q_L	Mdn.	Q_U and Q_L	Mdn.	Q_U and Q_L
Trials to criterion	16 trials	10.0	8.5, 11.5	10.5	7.0, 11.5	10.5	9.5, 12.5
Median response time (sec.) in 1st 9 trials	14 trials†	1.6	0.9, 1.9	1.8	1.2, 2.0	1.8	1.0, 2.2

* Difference between these two groups has a one-tailed p value of .02 for the response-time measure.

† Original data for two Ss in the no-shock group were lost.

The adult avoidance-training data appear in Table 3. The table presents two measures: (a) trials to achieve an avoidance criterion of five successive avoidances, including the criterion trials; and (b) median response time for the first nine trials since nine trials were the least any S received. According to both measures of performance, the response-contingent group was superior to both other groups. However, only the difference in running times between the two theoretically extreme groups, the response-contingent and the no-shock, achieved statistical significance according to Wilcoxon's (9) non-parametric test for grouped data. The grouped data rather than the paired replicates test was used because the adult avoidance training was carried out at various ages. The source of variation due to this variable plus that correlated with litters was eliminated in the statistical test by treating the data as replications of comparisons involving four homogeneous Ss. The quartiles which appear in Table 3 are included to indicate the rather small amount of variability present in each group, and thus the simplicity of the adult avoidance learning task.

DISCUSSION

The open-field data clearly provide no support for the hypothesis that infantile training with escapable shock leads to less fearfulness than infantile exposure to nonescapable shock. It might be argued that this negative finding follows from the fact that our infantile training procedure failed to demonstrate any learning to escape shock as measured by the performance of the infantile mice. A better test of the hypothesis would have been provided if our response-contingent infantile group had actually im-

proved in performance during their infantile training.

On the other hand, it should be recognized that our response-contingent shock offset procedure did assure that each time actual offset of shock occurred, the infantile S's last response to shock on each trial was locomotory movement. In this sense, the response-contingent group may be contrasted with the arbitrary shock group where such a consistent relationship between shock, locomotion, and shock offset was not assured, and even more clearly contrasted with the no-shock group. Thus, our finding that in adult avoidance training the more sensitive measure, running time, showed the response-contingent group reliably superior to the no-shock group is most suggestive of further research.

It could indicate: (a) that the specific S-R relationships between shock and locomotory movements were strengthened during the infantile training; (b) that the infantile training provided both the response-contingent and the arbitrary shock groups with a learned fear drive (3) which was not evinced in faster running during adulthood by the arbitrary shock group because for this group during infancy shock terminated on some trials while Ss were not moving, thus theoretically strengthening responses incompatible with locomotory movements; or (c) that the response-contingent group was generally most fearful of the three groups. To argue for either a or b would require the further assumption that the infantile training procedure was too insensitive to provide a good measure of learning since the response times of the two shock groups did not differ during the infantile training. To argue for c would require the assumption that excretory responding was

too insensitive as a measure of fearfulness. In this connection, it is worth mentioning that *E*'s impression was that the response-contingent *Ss* were the easiest to pick out of the "safe" compartment of the adult training apparatus.

However, if we take the statistically significant finding at face value and also take into account the fact that our open-field data provide no evidence for differential general fearfulness among our three groups, then our positive finding indicates a remarkable *specificity* in the effects of infantile exposure to shock on later behavior. The response-contingent group was favored by both stimulus specificity (i.e., shock was used in infantile and adult training) and response specificity (i.e., the infantile training correlated shock termination with locomotory movements and the adult training also required locomotory movements). The data clearly suggest that it may prove heuristic to conceptualize the relationship between infantile experience and adult performance in terms of such specific transfer of training.

The present findings are also relevant to the "critical period" hypothesis proposed by Scott and Marston (5). The hypothesis divides mammalian behavior into four periods, which are regarded as providing new and marked opportunities for relatively permanent behavioral integrations based on conditioning, socialization, and maturation of the nervous system. Scott and Marston suggest that conditioning should be difficult, or unstable if achieved at all, during the first or neonatal period. This period lasts until the mouse is about 12 days old. Thus, the superiority of the response-contingent group as compared with the no-shock group in the adult avoidance test is not in accord with the postulated instability of conditioning during the neonatal period.

SUMMARY

Mice, 8 to 11 days of age, were subjected to three types of infantile treatment: (a) stimulation with shock which *S* could terminate by crawling to the safe side of the apparatus (the response-contingent group); (b) stimulation with shock which was merely matched in dura-

tion to that received by *Ss* in the response-contingent group, with escape from the shock compartment barred (the arbitrary shock group); and (c) exposure to the shock apparatus for matched durations of time without shock being present (the no-shock group). An open-field test of emotionality conducted when *Ss* were 45 to 54 days old showed no significant differences in the excretory behavior of the three groups. An adult avoidance learning test conducted when *Ss* had a mean age of 90 days showed that the running times of the response-contingent group were slightly but statistically significantly better than those of the no-shock group. The present experiment provides no evidence for any general emotional or traumatizing effect of either type of infantile shock stimulation, but does indicate the presence of a specific behavioral effect which is most readily understandable in terms of the correlation between shock, locomotory movement, and shock termination in the infantile training of the response-contingent group.

REFERENCES

1. HALL, C. S., & WHITEMAN, P. H. The effects of infantile stimulation upon later emotional stability in the mouse. *J. comp. physiol. Psychol.*, 1951, **44**, 61-66.
2. LINDZEY, G. Emotionality and audiogenic seizure susceptibility in five inbred strains of mice. *J. comp. physiol. Psychol.*, 1951, **44**, 389-393.
3. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
4. MOWRER, G. H., & VIEK, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193-200.
5. SCOTT, J. P., & MARSTON, MARY-VESTA. Critical periods affecting the development of normal and mal-adjustive social behavior of puppies. *J. genet. Psychol.*, 1950, **77**, 25-60.
6. SNELL, G. D. *Biology of the laboratory mouse*. Philadelphia: Blakiston, 1941.
7. SOLOMON, R. L., & WYNNE, L. C. Traumatic avoidance learning: acquisition in normal dogs. *Psychol. Monogr.*, 1953, **67**, No. 4 (Whole No. 354).
8. WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.
9. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.

Received July 15, 1955.

PERFORMANCE OF FUNCTIONAL PSYCHOTICS ON A REPETITIVE TASK¹

P. H. VENABLES AND J. TIZARD

Medical Research Council, Unit for Research in Occupational Adaptation, Institute of Psychiatry, London, England

AN experiment by Kraepelin (3) involving a task of simple addition for periods of ten minutes demonstrated that the performance of schizophrenics is characterized by a rapid deterioration. When a pause is introduced after five minutes, performance increases markedly on the sixth minute, but then sinks to a low level during the remaining four minutes. Hoch (1) reported that manic-depressive patients show reversals of fatigue effects on the ergograph. Performance was found to increase in strength before a reduction appeared with exhaustion. In view of the somewhat unsatisfactory reporting of these experiments and the presence, particularly in the addition task, of interfering factors such as pauses at the end of a line of figures, it was decided to make an attempt to repeat, on an objectively scorable task, the work of these earlier writers.

Pilot experiments suggested that addition and color naming involving verbal reports were unsatisfactory because of practice effects on the addition test and pauses due to breathing in the color-naming test. The investigation to be reported was made using a five-choice task where scoring was entirely objective; no interruptions were made to disturb *S*'s output, and the task, involving less cerebration than does addition, was within the ability of *S*s showing a marked degree of psychosis.

METHOD

Apparatus

The apparatus was designed to provide a means whereby response to a light stimulus is instrumental in setting up a subsequent stimulus, this process being capable of indefinite repetition. The *S*'s panel consists of five lights and five keys arranged in pairs symmetrically around the circumference of two concentric semicircles. The inner circle on which the keys are placed is seven inches in diameter, the outer with the lights is ten inches in diameter. At the center of the semicircles is a sixth key. In operation, *S* presses the

key corresponding to a light which is lit and then the center key, whereupon the light changes. The sequence is continued by pressing the key next to the light which is then illuminated, and so on.

The control of the display is achieved by a uni-selector and relay mechanism modified from that designed by Leonard (4).

An event recorder recorded separately both correct responses and errors. A time marker permitted taking a count of responses in any period.

Subjects

Thirty male schizophrenics and 10 men suffering from endogenous depression were used as *S*s. They were selected after a scrutiny of hospital records and an inquiry to charge nurses as to whether they thought the patients would be willing to cooperate. Six other patients selected in this way did not complete the tasks. No *S* was over 55 years of age. Diagnoses had been made by the medical superintendent for official records.

The schizophrenics were divided into three groups by age (20-29, 30-39, 40-55 years), and two groups by length of stay. Short-stay patients had been in the hospital less than one year from their most recent admission; long-stay patients had been in the hospital for more than two years since their last admission. There were thus six groups of schizophrenics, with five *S*s in each group. All were diagnosed as suffering from other than paranoid types of schizophrenia (simple, catatonic, or unclassified) without organic or other complicating features. Paranoid schizophrenics were excluded because a pilot study had shown that their response patterns differ from those of other schizophrenics. All except one of the short-stay patients had had more than one admission to hospital. No *S* was receiving physical treatment or sedatives at the time of the investigation.

The 10 endogenous depressives had a mean age of 49.4 with an *SD* of 3.6 years. In age they were comparable with the 40-55-year-old schizophrenic group, who had a mean age of 47.3 and an *SD* of 2.9 years. Seven *S*s had had more than one admission to hospital with a depressive diagnosis; three involutional depressives were in hospital for the first time.

Procedure

The *S*s were told that the test was "to see how quick they were" and that any questions they asked would be answered at the end of the experiment. Each *S* was given a 10-minute run on the task followed by a rest pause of one minute and a further 10-minute run. During the rest pause, *S* was engaged in conversation.

A continuous record of responses was made for each *S*. To simplify the analysis, the scores used for further

¹ This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, Dr. H. C. Beccle. Advice on the statistical analysis was given by Mr. A. S. C. Ehrenberg.

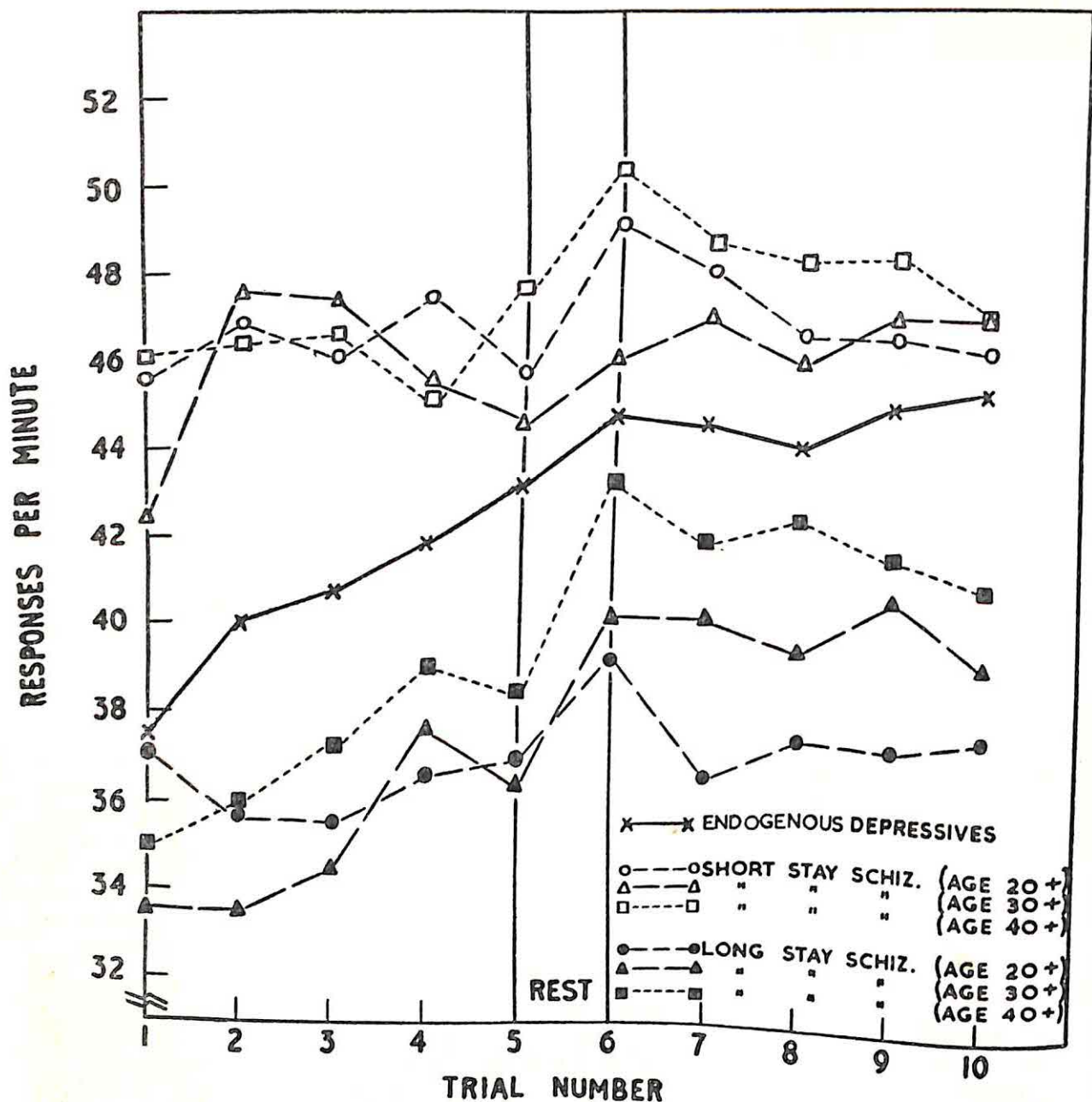


FIG. 1. MEAN SCORES PER MINUTE OF SCHIZOPHRENICS AND ENDOGENOUS DEPRESSIVES ON A REPETITIVE TASK. $N = 5$ IN EACH SCHIZOPHRENIC GROUP AND 10 FOR THE DEPRESSIVE GROUP

calculation were the average score per minute of successive two-minute periods. Linear regression constants, calculated for each S , summarized his scores and indicated his initial level and trend of performance in each 10-minute work period.

RESULTS

The mean scores per minute for each of the schizophrenic groups and for the depressives are shown graphically in Fig. 1. Application of analysis of variance to the regression con-

stants obtained from the schizophrenic groups gave the following results.

There were no significant differences among the schizophrenic groups in slope (trend of performance). All showed a tendency to rise during the first 10 minutes and to fall during the second ten minutes.

The differences in level (speed of performance) between the three age groups were not significant. There was, however, a difference

TABLE 1

ANALYSIS OF VARIANCE OF INITIAL LEVEL OF PERFORMANCE OF SCHIZOPHRENIC GROUPS IN PERIOD 1

Source	df	Mean Square	F
Age	2	11.2	0.15
Length of Stay	1	898.7	12.19*
Interaction	2	2.31	0.03
Within Cells	24	73.72	
Total	29		

* Significant at the .01 level.

TABLE 2

ANALYSIS OF VARIANCE OF INITIAL LEVEL OF PERFORMANCE OF SCHIZOPHRENIC GROUPS IN PERIOD 2

Source	df	Mean Square	F
Age	2	30.1	0.36
Length of Stay	1	444.67	5.31*
Interaction	2	12.54	0.15
Within Cells	24	83.81	
Total	29		

* Significant at the .05 level.

in level (significant at the .01 level in Period 1 and at the .05 level in Period 2), between short-stay and long-stay schizophrenic Ss, the short-stay Ss being superior to the long-stay Ss. Tables 1 and 2 show the analyses of variance of initial level of performance of the schizophrenic Ss in Periods 1 and 2.

An increase in speed of performance after rest was shown by 26 out of the 30 Ss who improved their scores when beginning Period 2. By the sign test, this increase is significant at the .01 level.

When a comparison was made between the depressives and the schizophrenics who resembled them most, the 10 Ss aged 40 and over, it was found that the performance of the two groups differed significantly in slope, the difference being significant at the .05 level for both Periods 1 and 2 ($t = 2.33$ in each case). The depressives rose steadily in performance throughout both Period 1 and Period 2, whereas the schizophrenics' performance rose much less in Period 1 and declined in Period 2.

The schizophrenic group improved considerably in performance after rest (a finding also reported by Kraepelin), whereas the depressives showed only a slight improvement.

"Improvement scores" were calculated for each patient by subtracting the last score of Period 1 from the first score of Period 2. Owing to the wide scatter of schizophrenic scores, there was no significant difference between the "improvement scores" of the two groups of Ss being compared when a t test was used ($t = 1.22$). However, the nonparametric median test gave a highly significant chi square between the two sets of scores ($\chi^2 = 8.18 = p < .01$). Since all six schizophrenic groups show a larger rise in level of performance after rest than do the depressives, the median test probably gives a valid indication of a true difference between the groups as far as increment in performance after rest is concerned.

DISCUSSION

The findings of Kraepelin of an immediate fall in the performance of schizophrenics have not been confirmed in this study, although improvement after a pause and subsequent deterioration in performance were shown. The initial rise in performance shown here is most likely due to the effect of practice on an unfamiliar task. That a pilot study preliminary to the one reported here showed marked practice effects on an addition task, contrary to Kraepelin, may be because his results are expressed as the means of performance over 10 days, during which time his subjects would become practiced on the task. Also, Kraepelin's Ss, details about whom are not given, were possibly of a higher educational level.

In spite of these differences, both the current findings and those of Hoch and Kraepelin can be subsumed under a theoretical framework derived from Pavlov and stated in Hullian (2) terms. Pavlov (5) stated that schizophrenics were characterized by a "weak" nervous system with a predominance of inhibitory potential, whereas depressives presumably had a predominance of excitatory potential.

In Hullian terms, schizophrenics might be said to show an exaggerated tendency to develop reactive inhibition in contrast to depressives. In Kraepelin's experiment, the fall in performance could thus be attributed to a rapid development of reactive inhibition in a familiar task. That is, the change in $s\bar{E}_R$ is predominantly due to the development of I_R . In the present results, the increasing $s\bar{E}_R$ in schizophrenic performance may be conceived

as a function of increasing sH_R . The sE_R is, however, depressed by a development of I_R . In comparison, the depressives, having a weaker tendency to develop reactive inhibition, showed a performance where sE_R was almost solely a function of increasing sH_R . In both studies, the increase in performance after rest could be due to the dissipation of I_R the rather more marked improvement with Kraepelin's subjects being due to the longer (five-minute) rest pause which he employed. Hoch's results also seem consistent with the interpretation of little development of reactive inhibition among depressives.

While the findings in these three studies are thus susceptible to a theory of differential development of reactive inhibition, other explanations are possible. Kraepelin himself dismissed the notion of fatigue as being responsible for the decrement in schizophrenic performance and favored an explanation in terms of diminishing "willed attention" (*Willenspannung*). The present finding of an initial improvement in performance would appear to run counter to his explanation.

The superiority in level of performance shown by the short-stay over the long-stay group of schizophrenics is a not unexpected finding which exemplifies the general slowness of response presented by most institutionalized patients. An explanation of this schizophrenic deterioration in terms of lack of drive (D) is possible, but like the theoretical interpretation presented earlier, must await further experimentation for its resolution.

SUMMARY

When given a repetitive task to perform for two successive periods of 10 minutes, non-paranoid schizophrenics tended to show little improvement in performance with practice during the first period, a rise in level of performance after rest, and a subsequent fall in level during the second period. Endogenous depressives, on the other hand, showed an improvement in performance throughout both periods, but little increase in level of performance after rest. Their performance differed significantly from that of the schizophrenics in these respects. Within the groups of schizophrenics, there were differences in response rate depending on length of stay in hospital.

The findings support results obtained by other workers and are consistent with a theoretical framework stated in terms of differential susceptibility to reactive inhibition.

REFERENCES

1. HOCH, A. On certain studies with the ergograph. *J. nerv. ment. Dis.*, 1901, **28**, 620-628.
2. HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
3. KRAEPELIN, E. *Psychiatrie* (Vol. 3). Leipzig: Barth, 1913.
4. LEONARD, J. A. Advance information in sensorimotor skills. *Quart. J. exp. Psychol.*, 1953, **5**, 141-149.
5. PAVLOV, I. P. *Conditioned reflexes and psychiatry*. (W. H. Gantt, Trans.). London: Lawrence & Wishart, 1941.

Received July 29, 1955.

THE LEARNING AND UTILIZATION OF CONTRAVALUANT MATERIAL¹

EDWARD E. JONES AND JANE ANESHANSEL

Duke University

DURING the past two decades, there have been several attempts to investigate the effects of attitude, set, and frame of reference on the learning of meaningful prose material. The general form of these studies has involved the comparison by various measures of immediate and delayed recall, of groups known or presumed to differ in their attitude toward the content of the material to be learned. In an early study, Laird (5) found that optimistic *Ss* recalled significantly more pleasant items than did pessimistic *Ss*. Widely separated studies by Clark (2) and Alper and Korchin (1) have both explored differences between the recall by males and females of controversial prose material presumably relevant to attitudes determined by sex identification. Both studies resulted in differences between recall scores of males and females which were interpretable in terms of the material presented. Watson and Hartmann (9) found that theistic and atheistic students tended to remember better material supporting their "attitudinal frame," although most of the differences in their study were not statistically significant. With a larger group of *Ss* and a more sensitive method of quantitative analysis, Edwards (4) came to the same general conclusion with regard to *Ss* differing in their attitude to the New Deal. The most familiar study in this vein is the clear-cut demonstration by Levine and Murphy (6) that prose which is congenial either to pro-communists or to anti-communists is more rapidly learned and more slowly forgotten than prose which presents arguments running counter to the *S's* attitudinal bias.

In general, considering the difficulties involved in the reliable assessment of relevant attitudes, the assumption that we more easily learn that which is congenial to our own views appears to have fairly solid empirical support. While previous authors have generally been cautious in generalizing their findings to other attitudes in other learning situations, it has become convenient to stress the autistic nature

of meaningful learning at the expense of other motivational effects which may operate to inhibit or actually reverse the outcome predicted by a simple autism theory. This paradox has been recognized by many of the students of social perception who have explored the polar mechanisms of perceptual defense and perceptual vigilance. To resolve this paradox, the functionalist might suggest that we examine the total context in which perception (or learning) takes place, in an effort to determine those conditions which promote lowered thresholds for threatening material (perceptual vigilance) and those which promote heightened thresholds (perceptual defense). On the basis of our analysis of these conditions we should be able to decide whether defense or vigilance will have greater functional utility in a given situation, and to predict the corresponding direction in threshold variation.

If we assume a continuity between perception, learning, and retention (and we would be in good company in doing so), there must be conditions under which the learning counterpart of perceptual vigilance will occur. It is somewhat surprising that as yet no one has thought to explore the conditions under which uncongenial or contravaluant material is successfully learned. It is rather easy to think of exceptions to the proposition that our minds are furnished largely with covaluant ideas. One need merely think of the political campaigner digging into his opponent's record, or the intelligence agent plotting the course of Soviet diplomacy, or the liberal who quotes *Time* magazine with shocked bitterness, to appreciate that we do not always surround ourselves with ideological yes-men. It is obviously an important psychological question, however, whether our pains to learn an opponent's views are rewarded with accurate learning and retention, and whether the conditions under which such learning is facilitated may be specified.

The present paper attempts to evaluate the effectiveness of a particular set of conditions in promoting the learning of contravaluant or uncongenial prose sentences. In brief outline, we propose to test the hypothesis that *Ss* who are presumed to disagree with a set of state-

¹ The authors are indebted to Drs. Gregory Kimble and Harry I. Kalish for their helpful suggestions and general facilitation of the research.

ments will actually learn them better than Ss who agree when retention of the statements serves a particular function in a later part of the experiment. Specifically, when Ss are told that they will be asked to provide appropriate counterarguments for prosegregation statements, prosegregation Ss will learn a set of antisegregation statements *better* than antisegregation Ss. In keeping with the traditional argument, however, antisegregation Ss should learn more antisegregation statements when there is no anticipation of a counterargument task.

METHOD

Subjects. A 10-item Likert-type scale was devised for the experiment as a measure of attitudes toward segregation. The scale was pretested on 151 introductory psychology students and one item was subsequently modified because it showed a poor correlation with total score. The final form of the scale was administered to 147 different introductory students and the scores ranged from 61 (out of a theoretically possible high of 70) to 15 (out of a possible low of 10). The distribution of scores was approximately normal for both groups. While the segregation questionnaires were filled out anonymously, the 20 highest and 20 lowest Ss were reliably identified by other data and were successfully approached to serve as Ss in the apparently unrelated experiment. The scores of the prosegregation Ss (hereafter called "pros") ranged from 61 to 38; the scores of the antisegregation Ss (hereafter called "antis") ranged from 15 to 22.

Segregation attitude scale. Each of the 10 items of the segregation scale was explicitly related to the segregation issue. Although the scale itself was unlabeled, no special pains were taken to disguise intent because of the nature of our sample. The origins of the Duke student body are geographically diverse (with slightly over half coming from states in which segregation is sanctioned and practiced), the issue of segregation is a lively one on the campus, and there is no clearly specified college norm which would promote the uniform denial of prosegregation statements. Five of the scale items were slanted in the prosegregation direction (e.g., "Any attempt to deal with segregation ignores the oft proven maxim that you can't legislate morals") and five were slanted in the antisegregation direction (e.g., "The Negroes' major concern is with equal educational and economic opportunities. They have no intentions of interfering with the social patterns of the white community").

The validity of the scale is one of the issues at stake in the experiment itself, of course, but there are two independent sources of evidence that suggest that the scale is sufficiently valid for the purposes of the study. The original pretest group of 151 Ss was simultaneously administered a combined version of the California Ethnocentrism and Anti-Semitism scales (with some additional pro-Semitic items included to break any response set). This combined scale—the majority of whose items deal with attitudes towards Jews—corre-

lated .41 with the segregation scale. Perhaps more revealing, in terms of validity, is the finding that the mean segregation-scale score of Northerners (29.7) differs significantly ($p < .01$) from the mean segregation score of Southerners (35.4). This contrast is in spite of the fact that the general prejudice scores of the two groups are indistinguishable (89.5 vs. 89.6). It would seem legitimate to assume that the Ss selected for the experiment itself had rather consistent beliefs either for or against segregation.

Experimental procedure. Each of the 40 Ss individually attempted to learn 11 brief statements which were clearly loaded in the direction of arguing against segregation. The statements were printed on individual cards, and each S was instructed to read each card aloud at normal reading speed. After all 11 statements had been read in this manner, the S was asked to reproduce as many statements as he could remember. This identical procedure was repeated for a total of five trials. The order of the statements differed randomly on each trial, but the order was constant across Ss. All attempts at reproduction were transferred verbatim to specially devised recording sheets in order to minimize recording error. In addition, the experimenter was usually unaware of whether a particular S was pro- or antisegregation.

After the five learning trials, each S was presented with a list of 11 prosegregation statements, and asked to provide appropriate counterarguments for each statement as rapidly as possible. Finally, each S was asked to match each pro- with each antisegregation statement, and told to indicate which of the statements came closer to his own beliefs.

While all of the Ss followed this same general procedure, half of the Ss (10 "pros" and 10 "antis") were forewarned that they would be asked to provide counterarguments for prosegregation statements. These Ss will be hereafter designated as the *experimental group*. Since the precise nature of the instructions to this group is important in evaluating the experiment, they are reproduced below:

"This is going to be an experiment to see how well you can think up appropriate counter-arguments for controversial statements. You will be presented with a number of statements which all argue more or less in favor of segregation. Your task will be to look at each statement in turn, read it aloud, and then give me an appropriate counter-argument. . . (They were then provided with an example of a statement and its counter-argument.) . . . Before getting on with the main part of the experiment, I am first going to show you some anti-segregation statements like the ones you may want to use as counter-arguments. Your first job is to learn *these* statements as quickly and completely as you can. Remember, you may want to use some of the statements in the main part of the experiment later on, so it will pay to learn these statements."

The other half of the Ss (the *control group*) were simply instructed to learn the antisegregation statements. No reference was made to the possible use of these statements in a subsequent part of the experiment. As indicated above, however, all Ss actually were eventually placed in the same position of using the statements as counterarguments.

The following examples illustrate the kind of statements which the Ss were originally asked to learn. Each

antisegregation statement is followed by its appropriate counterargument (as determined by the consensus of several graduate student judges). The counterargument, of course, only appeared along with other counterarguments during the second phase of the experiment.

1. The issue of Negro-White integrated education has nothing to do with racial intermarriage.

Counterargument: The Negro mulatto is a symbol of what would happen after desegregation.

2. The Negro points up the greatest disparity between the theory and our practice of democracy.

Counterargument: The democratic philosophy includes the right to associate with whomever you please.

No claims need be made for the logical or empirical validity of any of these statements since they were intentionally stated in somewhat controversial form.

RESULTS

The Learning of Controversial Statements

The major analysis involved scoring for accuracy each reproduction attempt for each *S* for each trial. The 11 antisegregation statements were broken down into 22 idea units and these in turn were scored for accuracy in terms of the following three categories: 1, the idea unit is identical to the original, or modified only in some incidental way (e.g., the *S* says "because of" whereas the original reads "a function of"); 2, the idea unit is identical in meaning to the original, though it may vary in phraseology, through the use of synonyms, etc. (e.g., the *S* says "integrating the Negroes with the whites in education" whereas the original reads "Negro-White integrated education"); 3, the idea unit clearly conveys a different meaning than was intended by the original. Ten randomly selected response protocols (consisting of 632 idea units) were independently scored by a second rater. There was 89 per cent agreement on item placement.

The major hypothesis to be tested is that the Experimental "pros" will learn more idea units than the Experimental "antis," and the Control "antis" will learn more idea units than the Control "pros." In order to test this hypothesis, units learned were placed in a 2 (for replicated *Ss*) by 2 (for conditions) by 5 (for trials) factorial design and, following Lindquist (7), a Type III analysis of variance was conducted. Three separate analyses were made, covering respectively the number of units exactly reproduced (Scoring Category 1), those which were reproduced with unchanged meaning (Category 1 + 2), and total units tried (1 + 2 + 3). These three analyses were con-

ducted in order to test the hypothesis under investigation at different levels of accuracy.

Turning first to the analysis based on the very stringent accuracy criterion of Category 1, the over-all means (combining trials) indicate a slight tendency toward the predicted interaction but the *F* ratio is far from significant. As Table 1 shows, however, there is a significant triple interaction between *Ss*, trials, and conditions. Figure 1 reveals that this is a function of clear-cut fluctuations in the relations between the four cell means from trial to trial. Trials 3 and 5 provide the clearest example of the predicted interaction between *Ss* and conditions, but separate single analyses of variance for these trials result in moderate but nonsignificant interactions.

When Category 1 is combined with Category 2 to provide a still meaningful but less stringent measure of accuracy of reproduction, the analysis of variance provides positive confirmation of the main hypothesis. As Table 1 and Fig. 1 show, at every trial Experimental "pros" learn more antisegregation units than Experimental "antis," and Control "pros" learn less than Control "antis." Except for the marked improvement in score as a function of trials, no other interaction or main effect is significant, and the application of Bartlett's test shows

TABLE 1
LEARNING AT THREE LEVELS OF ACCURACY
(“Pros” vs. “antis” under experimental and control conditions, analysis of variance)

Source	df	Stringency of Accuracy Criterion					
		Score 1		Score 1 + 2		Score 1 + 2 + 3 (total tried)	
		Mean Square	<i>F</i>	Mean Square	<i>F</i>	Mean Square	<i>F</i>
Between <i>Ss</i>	39						
A: Conditions	1	9.68	—	19.84	—	30.42	—
B: Attitude	1	.02	—	.24	—	1.62	—
A × B	1	52.00	1.10	315.01	6.39*	233.28	6.42*
Error	36	47.34		49.34		36.35	
Within	160						
C: Trials	4	465.09	112.48**	687.49	214.18**	674.28	156.93**
A × C	4	.84	—	1.73	—	.66	—
B × C	4	2.34	—	1.01	—	.63	—
A × B × C	4	12.45	3.01*	5.12	1.60	1.42	—
Error	144	4.13		3.21		4.30	
Total	199						

* *p* < .05.

** *p* < .01.

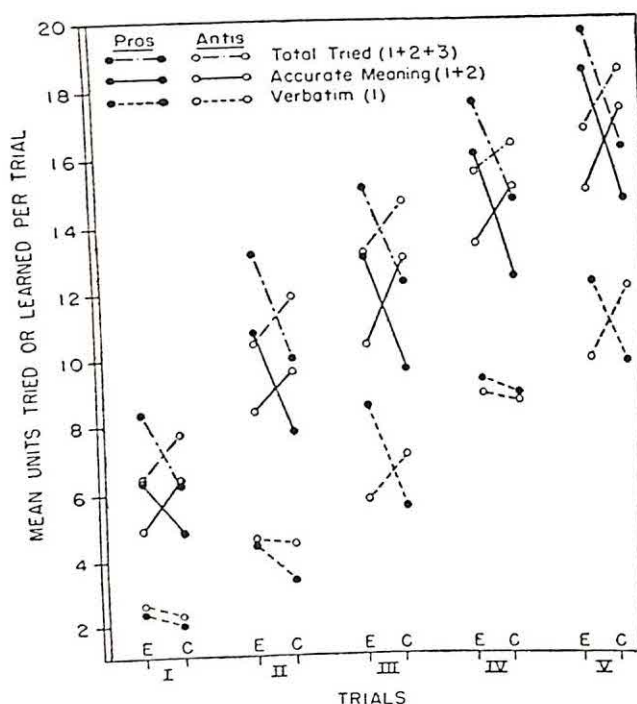


FIG. 1. MEAN UNITS TRIED OR LEARNED AT TWO LEVELS OF ACCURACY PER TRIAL. PROSEGREGATIONISTS VS. ANTISEGREGATIONISTS UNDER EXPERIMENTAL (E) AND CONTROL (C) CONDITIONS

that the variances in each cell are homogeneous. Within conditions, *t* tests reveal a significant ($p < .05$) difference between "pros" and "antis" under experimental but not under control conditions.

Since the second analysis of variance resulted in a significant interaction, it is only of incidental importance that the third analysis of variance (dealing with the total number of units tried) resulted in almost identical *F* ratios for all main effects and interactions, and the main effect of trials plus the conditions by *Ss* interactions are again the only significant sources of variance. A glance at Fig. 1 tells us that *Ss* seldom responded incorrectly if they responded at all, and the fact that the slopes indicating the shift in mean as a function of conditions are uniformly parallel suggests that nothing would have been gained by devising a ratio score reflecting the percentage of accurate reproductions out of the total tried. In other words, reproduction errors were apparently distributed randomly across trials, conditions, and *S* groups.

In general, then, the analyses offer strong support for the main hypothesis of the study if

we use a "meaningful" rather than a strictly verbatim criterion of accuracy. One apparent reason why the predicted interaction is significant under the lenient but not the stringent criterion of accuracy is that summing Categories 1 and 2 naturally results in a greater range of scores than Category 1 alone, thus providing a more reliable quantitative estimate of recall. However, a more purely psychological factor may also be operating to attenuate the predicted interaction under stringent accuracy conditions. Doob (3) found that whereas *Ss* tended to recall an equivalent number of congruent vs. incongruent paragraphs, a percentage measure of recall accuracy based on individual statements showed that items with which the *S* agreed were recalled more correctly than incongruent statements. Thus it may be that strict accuracy of recall is determined largely by one's attitude toward the material and is not as subject to modification by variation in motivating instructions as the total amount learned as judged by less stringent criteria. Nevertheless, it should be pointed out that the more lenient (1 + 2) measure of accuracy, which results in confirmation of the prediction in the present case, is very much a measure of "effective accuracy" since the meaning of the statement had to be preserved without modification to receive a score of 2.

In any event, conditions may apparently be established which can effectively reverse the customary conclusion that *Ss* will show better learning of congenial than uncongenial material. It is remotely possible that the present results reflect sampling errors in recruiting 10 *Ss* for each cell. It may be, for example, that the "pros" who ended up in the Experimental group were more intelligent, or more rapid and retentive readers than the Control "pros." This would not appear to be the case, if we can place any confidence in Total ACE scores as a measure of intelligence. While the "pros" do differ reliably from the "antis" in Total ACE (the difference between scores favoring the "antis," $p < .05$), there are no such differences in reading skill, or in either reading or total scores of the experimental versus the control "pros."

The Utilization of Statements as Counterarguments

The primary purpose for presenting pro-segregation statements to the *Ss* and eliciting appropriate counterarguments was to fulfill the

conditions of the experiment for the Ss. However, we thought it was possible that "pros" and "antis" would differ in the appropriateness of the argument chosen, and that perhaps the "antis" would show a less heavy reliance on the antisegregation statements learned. A scoring system similar to that used on the learning data was applied to the counterarguments chosen by the Ss. Five instead of three scoring categories were devised in an attempt to gain a more sensitive measure of variations in the kind of statement chosen as a counterargument:

1. entire statement essentially correct;
2. in spite of changes in phrasing, the meaning of the statement is correct;
3. the statement is appropriate but novel or creative;
4. an earlier learned statement is inappropriately applied;
5. nonsense statements and garbled versions are produced.

A series of 2×2 analyses of variance were conducted for each category and for various meaningful combinations of them, and the results are inconclusive. In eight separate analyses, not a single main effect or interaction was significant. There is a consistent, but never significant, tendency for the antisegregationists to be more "creative" under experimental conditions. Aside from this, there is little consistency in these data.

When the Ss were permitted to examine both lists of statements simultaneously and asked to match appropriate counterarguments, there were only a few "errors" in the entire sample of 40 Ss. The fact that 24 out of 40 Ss had completely correct matches supports the validity of our a priori selection of meaningful opposites. Perhaps because of the over-all high level of accuracy, there were no significant group differences in matching errors.

The final task for the Ss, it will be remembered, involved circling the alternative in each pair of counterarguments which best reflected their attitude. It should be noted that whether or not a "pro" indicates a preference for pro-segregation statements is a relevant but far from decisive test of the validity of the segregation scale and the appropriateness of the statements to be learned. The task of circling the preferred statement was completed under conditions of surveillance and nonanonymity, and furthermore, the statements were generally stated in an extreme form, so a clear-cut reflec-

tion of true attitude would not be expected. It is not surprising, therefore, that the modal number of Ss circled 9 antisegregation statements and 2 pro-segregation statements. With the exception of the 9 out of 40 Ss in this modal group, however, there is practically no overlap between "antis" and "pros." Application of the nonparametric Median Test results in a highly significant difference between pro-segregation and antisegregation Ss in the relative tendency to check pro-segregation statements. The fact that this difference appeared is encouraging evidence for the validity of the attitude scale and also for our assumption that the statements would be relatively more congenial to one group than to the other.

DISCUSSION

The results of the present experiment indicate that under certain conditions, Ss who disagree with a set of statements will actually learn them better than Ss who agree with the statements. This conclusion complicates the picture presented by those who favor consistent explanations in terms of need-relevant perceptual and mnemonic selectivity. The traditional interpretations of meaningful learning which stress autism, perceptual selectivity, or congruence with the learner's frame of reference, receive support under the standard control conditions of the present experiment, but a rather slight modification in experimental instructions is sufficient to reverse the conventionally predicted relationship. It would appear, then, that we tend to be autistic in the retention of congenial material under conditions of low motivation and circumscribed time span. When the stakes are raised and the context of behavior is enlarged by increasing the instrumental significance of learning contravaluant material, something like vigilance appears. For the pro-segregationist under the experimental conditions, the comfort of surrounding himself with congenial ideas is pitted against the comfort of performing well the stated experimental task. In this particular instance the latter factor won out. When the contravaluant material is given a new utility for the learner, the wish that is otherwise autistically pampered is supplanted by more powerful motives of ego-enhancement (subsuming implicit competition with other Ss as well as the hope of a kind word of reward from the experimenter). While this interpretation

supplements the restricted theories of autism and covalent selectivity, it neither supplants nor contradicts such theories. In particular, the present results support the functionalistic or pragmatic notion that we learn best that which is instrumental to adaptation and survival.

The interpretation outlined above would hold equally well if the predicted differences observed between "pros" and "antis" in the control group were merely eliminated under the experimental instructions. But why should the "pros" actually surpass the "antis" in the learning of antisegregation material? The clue to this paradox, we suggest, lies in the general contention that different individuals can be differentially motivated by the same set of instructions. We assume that both "pros" and "antis" are about equally motivated to do well in the experiment—i.e., they both want to prepare themselves for the task of countering pro-segregation arguments with appropriate anti-segregation statements. While the "antis" presumably come trained and primed for such a task, the "pros" face the prospect of adducing contravalent arguments like a nervous understudy rushed on stage to play an unfamiliar character type. Since the "anti" comes to the experiment more confidently familiar with the arguments which favor his cause, he is less inclined to view the initial learning task as a critical prerequisite for the fulfillment of the main experimental requirement. The "pro," on the other hand, is forced to rely on rote learning in order to comply with the experimental task and therefore he does a better job in memorizing the presented antisegregation statements. There is some empirical support for this otherwise plausible interpretation in the fact that the "antis" did tend to be more "creative" in their final rebuttals than the "pros," but this tendency was neither striking nor statistically reliable.

If we had introduced a third condition in which an extrinsic monetary reward was promised for successful learning, the same results would not be predicted. Assuming that the proposed monetary reinforcement was sufficient to raise the level of motivation to a high level in all Ss, we would predict an absence of any differences in learning between the two attitudinal groups—or rather, those differences that did exist would be a function of ability ceiling and not attitude. The reversal effect obtained

in the present experiment is dependent on more subtle reinforcement potentials intrinsically relevant to the task at hand. What we apparently have in the present case is the operation of self-instructions or covert sets (8) which involve different interpretations of the overt instructions. The "pros" and the "antis" are both motivated to attain the same goal, but they tell themselves different ways to achieve it.

SUMMARY

The assumption that we tend to learn statements that we agree with better than statements with which we disagree must be placed in the broader context of the learner's over-all purpose in the task. An experiment was conducted in which pro-segregation and antisegregation Ss were asked to learn a series of antisegregation statements. Half of the Ss proceeded under the assumption that these same statements could well be used as counterarguments for pro-segregation statements to be presented in a subsequent portion of the experiment. The remainder of the Ss were simply asked to learn the statements as well as possible. The results demonstrate quite conclusively that pro-segregationists will learn antisegregation statements *better* than antisegregationists when a subsequent debate is anticipated. Without such an anticipation, however, the customary finding holds—i.e., the antisegregationists make better progress than the pro-segregationists in learning the congenial material. No significant differences were found with regard to the utilization of antisegregation counterarguments in the debating task. The results were interpreted in terms of the differential availability of means to complete the experimental task. Since the antisegregationists are initially better prepared to fulfill the ultimate experimental requirement, they are less highly motivated to learn the congenial statements when a debate is anticipated.

REFERENCES

1. ALPER, THELMA G., & KORCHIN, S. J. Memory for socially relevant material. *J. abnorm. soc. Psychol.*, 1952, **47**, 25-37.
2. CLARK, K. B. Some factors influencing the remembering of prose material. *Arch. Psychol.*, N. Y., 1940, **36**, No. 253.
3. DOOB, L. W. Effects of initial serial position and attitude upon recall under conditions of low motivation. *J. abnorm. soc. Psychol.*, 1953, **48**, 199-205.

4. EDWARDS, A. L. Political frames of reference as a factor influencing recognition. *J. abnorm. soc. Psychol.*, 1941, **36**, 34-50.
5. LAIRD, D. A. The influence of likes and dislikes on memory as related to personality. *J. exp. Psychol.*, 1923, **6**, 294-303.
6. LEVINE, J. M., & MURPHY, G. The learning and forgetting of controversial material. *J. abnorm. soc. Psychol.*, 1943, **38**, 507-515.
7. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
8. POSTMAN, L., & SENDERS, VIRGINIA L. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, **36**, 153-165.
9. WATSON, W. S., & HARTMANN, G. W. The rigidity of a basic attitudinal frame. *J. abnorm. soc. Psychol.*, 1939, **34**, 314-335.

Received August 8, 1955.

AUTHORITARIANISM AND REPRESSION¹

NATHAN KOGAN

Harvard University

RECENT years have witnessed an increasing number of studies experimentally confirming many of the propositions set forth in *The Authoritarian Personality* (1). With few exceptions, investigators have focused upon the cognitive styles of ethnocentric and authoritarian individuals (2, 3, 4, 7, 8). Generally overlooked is the tendency to repress, one of the most basic features asserted to be typical of authoritarian personalities. According to Adorno, *et al.* (1), the highly authoritarian individual does not become aware of undesirable or hostile feelings toward parental figures. Feelings of inner weakness or dependency are similarly denied. The nonauthoritarian individual, according to these authors, is more likely to be an intellectualizer. He freely admits into awareness ambivalent feelings about parental figures and, in addition, conflicts in the sexual and aggressive areas are subject to conscious reflection with consequent guilt feelings.

Hyman and Sheatsley (5) in a methodological critique of *The Authoritarian Personality* have cast considerable doubt upon the data from which the presence of repression was inferred. An experimental examination of the relationship between repression and authoritarian-ethnocentric attitudes is, therefore, of particular relevance at the present time. Given the postulated link between repression and authoritarianism, the hypothesis is tested that recognition accuracy for stimuli consisting of disparaging aggressive and sexual references to parents and self is inversely related to the strength of authoritarian and ethnocentric attitudes.

¹ The present research is based upon portions of a Ph.D. thesis submitted to the Department of Social Relations, Harvard University. I am especially indebted to Professors G. W. Allport and J. S. Bruner for their invaluable suggestions and criticisms. Dr. R. Tagiuri read the manuscript and contributed many helpful comments. I am grateful for the assistance rendered by Mr. M. B. Schlank, who served as the independent judge. This research was facilitated by a Sigmund Livingston Fellowship awarded under the auspices of the Anti-Defamation League.

METHOD

In testing the hypothesis a technique first employed by Lazarus, Eriksen, and Fonda (6) was used. Forty-five sentences partially masked by white noise were recorded on tape. The first three sentences, all emotionally neutral in content, were practice items. There were 42 experimental sentences—14 neutral, 14 sexual, and 14 aggressive. Sexual and aggressive content was either directed against the self in the form of expressions of guilt and personal inadequacy, or against parents or people in general. Neutral sentences were included to take into account individual differences in auditory acuity. The three major types of sentences were randomized in presentation. Two examples of each—neutral, aggressive, and sexual, respectively—are presented below.²

"You like to go swimming during the summer season.

"The thought of a good meal increases your appetite.

"Only your death would be a just punishment for you.

"Your mother is to blame for your worst faults.

"You have been unable to break your ugly sex habits.

"You would have loved to share your father's sex life."

The auditory perception test was group administered. A tape recorder was situated in one corner of the experimental room and connected to a speaker placed in a prominent position approximately equidistant from all Ss. The instructions were recorded on tape and played over the speaker prior to the presentation of the first sentence. These instructions were as follows:

"You are now hearing my voice as recorded on a tape recorder. You should be able to hear me quite clearly. In just a moment I shall introduce a noise background along with my voice, and the experiment will begin. Do not hesitate to write down everything you hear regardless of how shocking or improbable it may sound. Even though you may not hear the entire sentence, try to make a meaningful sentence out of what you do hear. Guess when you are not sure. We will now begin."

The Ss were encouraged to make meaningful sentences to reduce the prospect of nonsense responses. Warning Ss of the shocking nature of the sentences was intended to discourage withholding of written responses to accurately perceived stimuli.

The first three sentences were practice items. The E presented each of these sentences, with Ss writing down what they heard. Then, E read the sentence aloud in its original form and replayed it again on tape (with noise background). The Ss could recognize

² A copy of the complete sentence list can be obtained from the author.

the sentence despite the noise after having been informed of its content. All Ss were thus convinced that they were not hearing meaningless sounds.

The 42 experimental sentences followed, the recording being stopped after each sentence to allow Ss to write down what they heard. The Ss were constantly cautioned not to look at each other's papers. Strict conformity to this rule was urged, since checking of responses among Ss would, in view of the emotional quality of the stimuli, obviously lead to inhibition of response.

All sentences received a quantitative score ranging from 0 to 4. The rating scale is described below:

4—perfect recognition.

3—almost perfect recognition, minor errors or omissions.

2—basic meaning of sentence recognized, with distortions or omissions in wording.

1—meaning of sentence not recognized, crucial omissions or distortions in wording, though at least half of the words in the sentence are used.

0—a blank, or mere stub of sentence perceived correctly; also a completely irrelevant response.

The individual sentence ratings were summated for the three types of sentences. To equalize the auditory acuity factor, a sex/neutral and an aggressive/neutral ratio was computed for each S.

The investigator scored all of the sentences for all Ss. To guard against the possible influence of rating bias, the sentence protocols were scored before the determination of the attitude scores. An advanced graduate student in clinical psychology also scored all of the sentences for Ss in the upper and lower quarters on the E and F scales which had been previously administered (see below.). The scoring key above was given to the judge, and illustrative examples were provided for each of the scoring categories. In no case did an illustrative example correspond to a sentence response in the data. The sex/neutral and aggressive/neutral ratios yielded Spearman rank-order correlation coefficients of .90 and .93, respectively, between the investigator and the independent judge.

The range of individual differences in auditory acuity for the present type of stimulus material would seem to be quite wide. Pretest Ss were able correctly to perceive, on the average, about one out of every three sentences presented. A small number, however, obtained a zero score, i.e., perceived none or very few of the words in any of the sentences. In the course of assigning quantitative scores to the experimental Ss, it became apparent that a considerable number of Ss failed to recognize enough of the stimulus material to yield a protocol of any value for quantitative scoring. A score of 10 assigned by each judge was therefore established as a minimum for the retention of a case. It should be noted that the reliability coefficients reported above are based upon Ss whose scores are above this arbitrary cutting point.

Consider next the middle range Ss whose sentence protocols were scored only by the investigator. Again, a score of 10 served as a cutting point, but the selection or elimination of an S rested exclusively upon the investigator's judgment. No great risk was involved, however, in view of the extent of agreement between the judges for upper and lower quartile Ss. Thus, the

mean score for retained Ss was 30.5 for the investigator and 31.6 for the independent judge. The mean difference yields an *SD* of 2.8. Since the lowest acceptable score among middle range Ss was 16, it is highly unlikely that any of them would have been eliminated had their sentence protocols been scored by the independent judge.

Of the 80 Ss—white Christian male freshmen at a large New England university—who participated in the perception study, 37 satisfied the criterion of recognition. The present report is based upon data yielded by these 37 Ss.

The Ss had filled out a 30-item attitude battery prior to their participation in the perception experiment. The California E and F scales each furnished 12 items.³ Since all of these items are worded in the same direction—agreement with the item is scored as authoritarian or ethnocentric—five items of opposite tendency were included to avoid the possible development of a negative response set in the low group. They were not scored.

The Ss could assign one of six ratings to each item ranging from +3 (very strong agreement) to -3 (very strong disagreement). These raw values were converted into scores ranging from 1 to 7, with a score of 4 assigned to the rare case of an omitted item. An S's score on E and F was obtained by computing the average of the converted ratings assigned to the twelve items selected from the respective scales.

RESULTS

Product moment r 's⁴ are -.39 between F scores and sex/neutral ratios and -.38

³ The writer wishes to thank Dr. D. J. Levinson for his assistance in the selection of items. Timeliness and satisfactory Discriminatory Power values dictated the choice of items. In addition, there was at least one item representing each aspect or subsyndrome of ethnocentric and authoritarian ideology. Eleven E-Scale items were taken from Table 19, p. 142 of *The Authoritarian Personality*—1, 2, and 3 of subscale A; 1, 2, and 6 of subscale B; and 2, 4, 6, 7, and 8 of subscale C. A sixth "patriotism" item—"The first thing foreigners in this country should learn is the true American way of thinking and acting"—was substituted for less timely items in subscale C to make a 12-item scale. Eleven F-Scale items were selected from Forms 45 and 40, pp. 255-257—6, 8, 9, 13, 18, 21, 25, 26, 34, 37, and 42. The twelfth item, a revision of Item 32 in Form 78, p. 226, rounded out the scale: "The best teacher or boss is the one who tells us just exactly what is to be done and how to go about it." Both new items have yielded satisfactory Discriminatory Power values in administrations of the E and F scales subsequent to the publication of *The Authoritarian Personality*.

⁴ Product moment r 's are justified in the present case, for the sample distributions of ratios were essentially normal. The mathematical property of ratios is that they approach zero as the numerator decreases and approach infinity as the denominator decreases. Hence, a markedly skewed population distribution would be the statistical expectation. Since a minority

TABLE 1
MEAN RECOGNITION ACCURACY SCORES FOR Ss IN
THE EXTREME QUARTERS OF THE F-SCORE
DISTRIBUTION

Sentence Type	Judge A*		Judge B*	
	Lower Q	Upper Q	Lower Q	Upper Q
Neutral	.95	.88	.97	.95
Sexual	.82	.37	.82	.43
Aggressive	.86	.46	.75	.46

Note.— $N = 11$ for both upper and lower quarters.

* Judge A is the investigator and Judge B is the independent judge.

between F scores and aggressive/neutral ratios. These values are significant beyond the .01 and .02 levels (one-tailed test), respectively. The correlation coefficients are negative since stronger authoritarian attitudes are associated with lower ratios, i.e., less accurate recognition of the emotional sentences. Product moment r 's between E scores and sex/neutral and aggressive/neutral ratios are $-.25$ and $-.09$, respectively. Neither of these values achieves statistical significance. The discrepant results for the two scales receive further consideration below.

One possible interpretation for the superior perceptual performance of Ss scoring low on the F scale is that these individuals are more intelligent and thus more likely to recognize the unusual content of the emotional sentences. Fortunately, a score for "Language" IQ derived from the California Test of Mental Maturity was available for 31 of the 37 Ss. When this variable is held constant, the partial correlation with F score is $-.32$ for sex/neutral ratios and $-.35$ for aggressive/neutral ratios. These values are significant at the .05 level (one-tailed test) in spite of the smaller number of cases. Thus, differences in perceptual performance cannot be attributed to differences in verbal aptitude.

Of the 37 Ss satisfying the criterion of recognition, eleven were in the upper quarter and eleven in the lower quarter of the F-score distribution. This finding indicates the lack of any basic acuity difference between the more and less authoritarian. Table 1 supports this

of Ss, however, showed superior recognition for the emotional stimuli, and then only to a very slight degree, the sample skewness is reduced drastically from what would be expected if ratios were formed by drawing random numbers.

contention. At auditory levels at which both upper and lower quarter Ss are able to identify neutral material, only the lower quarter Ss seem able to approximate this performance with the more emotional stimuli. The results for both judges achieve a high degree of agreement.

A comparison of the recognition accuracy for sexual and aggressive stimuli of individuals in the upper and lower quarter of the F distribution yields critical ratios with one-tailed probability values less than .05, based on the nonparametric Wilcoxon T test (10) applied to the sex/neutral and aggressive/neutral ratios. This level of significance was obtained when independent analyses were performed on data scored by the writer and by the independent judge. A comparable analysis of extremes on the E scale yielded nonsignificant critical ratios for both judges. These results corroborate the outcome of the correlational analysis and also serve as an indirect check on interjudge reliability.

DISCUSSION

The results clearly indicate the presence of an inverse relationship between authoritarian attitudes and the capacity to identify highly emotional stimulus sentences. While both high and low authoritarians perceive the emotional stimuli less accurately than the neutral material, the discrepancy is considerably greater in the case of the more authoritarian individuals.

Adorno, *et al.* (1) describe authoritarian individuals as repressors. According to these authors, such individuals demonstrate a lack of conscious ambivalence of a hostile or sexual nature toward parental figures and other persons (outgroup members excluded) and deny any manifestations of weakness or passivity relative to the self. If it is assumed that repression has perceptual consequences, the marked inability of the authoritarians to identify the sexual and aggressive sentences is experimental evidence in favor of the hypothesized relationship between repression and authoritarianism.

The hypothesis relating nonauthoritarianism and intellectualization is confronted with equivocal evidence in the present experimental findings. The nonauthoritarian Ss do not demonstrate superior recognition accuracy for the emotional stimuli when neutral sentences

serve as a baseline. In this sense, these Ss can be distinguished from Lazarus, Eriksen, and Fonda's (6) intellectualizing neurotics who perceived emotional stimuli more accurately than neutral stimuli. Of course, the latter set of results may be applicable only to pathological forms of intellectualization. White (9) in describing intellectualizing processes in obsessive neurotics refers to the detachment of sexual and aggressive impulses from the ego. This observation is in sharp contrast to the California investigators' description of the "ego-assimilated" nature of sexual and aggressive impulses in nonauthoritarians. Whereas intellectualization, in their conception, is an attempt to arrive at a genuine understanding of one's problems, neurotic intellectualization implies a shift of one's problems from a personal-emotional to an abstract impersonal plane.

This marked disparity between pathological intellectualization and the ego-defenses of nonauthoritarians calls into question the applicability of Lazarus, Eriksen, and Fonda's theoretical interpretation to the present findings. These investigators maintained that the emotional stimuli were equally threatening to their repressors and intellectualizers, but were perceived differently because of the particular ego-defense employed. In the present case, it seems much more probable that the distinctive personalities of authoritarians and nonauthoritarians make for differences in the subjective acceptability or "threateningness" of the stimuli. Nonauthoritarians who are allegedly capable of accepting their own sexual and aggressive impulses should perceive stimuli embodying such impulses as less threatening.

The degree of confidence that attaches to a repression interpretation would be considerably increased if an independent definition of defense mechanism were available. The lack of such a definition necessarily implies that the empirical findings do not provide a conclusive test of the hypothesized relationship between repression and authoritarianism.

The lack of positive results for the E scale deserves further consideration. Since the E scale measures ethnocentrism and prejudice, it is obviously less closely related to the authoritarian syndrome than is the F scale. Correlations between E and F are generally in the vicinity of .6 to .7. Given the fact that repression is explicitly an hypothesized component of "the authoritarian personality" and

only indirectly a component of ethnocentrism and prejudice, the discrepancy between E and F in the extent to which each is correlated with perceptual performance is not surprising.

SUMMARY

In an examination of the relationship between repression and authoritarian and ethnocentric attitudes, a group of male freshmen college students filled out the California E and F scales and then participated in an auditory perception experiment in which they were asked to identify a series of noise-masked self-disparaging sexual and aggressive sentences and neutral control sentences. The F scores were significantly related to the perception of sentences in both the sexual and aggressive areas with the more authoritarian men demonstrating poorer recognition accuracy. The E scale, on the other hand, did not prove to be related to perceptual performance. The interpretation of the experimental results stressed the greater "threateningness" of the stimuli for the authoritarians. Differences in perceived threat were, in turn, attributed to the manifestation of stronger repressive tendencies in the authoritarian Ss.

REFERENCES

1. ADORNO, T., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BLOCK, J., & BLOCK, JEANNE. An investigation of the relationship between intolerance of ambiguity and ethnocentrism. *J. Pers.*, 1951, **19**, 303-311.
3. FISHER, J. The memory process and certain psychosocial attitudes with special reference to the law of Pragnanz. *J. Pers.*, 1951, **19**, 406-420.
4. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108-143.
5. HYMAN, H. H., & SHEATSLEY, P. B. "The authoritarian personality"—A methodological critique. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality"*. Glencoe, Ill.: Free Press, 1954.
6. LAZARUS, R. S., ERIKSEN, C. W., & FONDA, C. P. Personality dynamics and auditory perceptual recognition. *J. Pers.*, 1951, **19**, 471-482.
7. O'CONNOR, PATRICIA. Ethnocentrism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526-530.
8. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.
9. WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.
10. WILCOXON, F. Individual comparisons by ranking methods. *Biometrics Bull.*, 1945, **1**, 80-82.

Received August 12, 1955.

THE AFFILIATION MOTIVE AND PERCEPTUAL SENSITIVITY TO FACES

JOHN W. ATKINSON¹ AND EDWARD L. WALKER

University of Michigan

THE present investigation is an attempt to predict from imaginative measures of the affiliation motive (*n* Affiliation) the frequency with which a subject (*S*) selects human faces from similar but nonhuman figures in a perceptual task. It was undertaken with several objectives in mind. The sensitizing effect of motives on perception has been demonstrated in numerous studies (7), but in many instances the results have admitted reasonable alternative interpretations. For example, McClelland and Liberman (6; 5, p. 257) have reported a study in which *n* Achievement scores obtained from imaginative (TAT) stories proved to be related to recognition thresholds for achievement-related words. Farber (4), in a careful analysis of the role of motivation in verbal learning and performance, points out quite correctly that these results can be interpreted as a function of verbal habits common to the imaginative story performance and the word-threshold measuring situation. It was hoped that in the present study we could examine the relationship between a verbal-imaginative measure of motivation and sensitivity to relevant stimuli in perception in a situation which would not admit Farber's alternative interpretation or explanation in terms of the word-frequency variable discussed by Solomon and Howes (8).

The second objective was to obtain further evidence of the validity of the method of scoring imaginative (TAT) stories designed to measure motivation for social acceptance or *n* Affiliation (2). The demonstration of a positive relationship between *n*-Affiliation scores obtained from imaginative stories and perceptual sensitivity to stimuli related to affiliation would be further evidence that a motivational disposition is being measured.

The senior investigator has discussed elsewhere (1, p. 90) the theoretical basis for an

expectation that when the motive is measured in a neutral situation (i.e., one in which no special operation has been performed either to arouse the motive or to create an especially relaxed condition), a relationship can be expected with dependent perceptual and performance variables only when arousal operations are performed immediately before the *S* begins the criterion task. A recent study by Moulton *et al.* (1) has found a clear inverse relationship between *n*-Achievement scores and recognition thresholds for success and failure words when the motive was aroused before the dependent variable performance, but a similar relationship of lesser magnitude with neutral conditions preceding the perceptual task. The difference between the two conditions was statistically significant, however, only in the case of failure words.

In the experiments to be reported, the strength of the affiliation motive was measured by having several groups of male college students write imaginative stories under normal (neutral) classroom conditions. These groups were subsequently subjected to procedures designed to measure perceptual sensitivity to affiliation-related stimuli. Some groups performed the perceptual task under affiliation-neutral conditions; other groups were exposed to procedures designed to arouse the affiliation motive prior to the measurement of perceptual sensitivity for affiliation-related stimuli.

METHOD

Subjects. The *Ss* were 93 male undergraduates who at the time of their participation in the experiments were members of an introductory experimental psychology course at the University of Michigan. They constituted nearly the entire population of men in three semesters of this course in the years 1952, 1954, and 1955.

Measurement of motive strength. In each class from four to six pictures were used to elicit imaginative stories which were scored for affiliation-related imagery to obtain measures of individual differences in strength of *n* Affiliation. The instructions used and evidence of the validity of the scoring system may be

¹ This research was conducted during the period of a Social Science Research Council Faculty Research Fellowship.

found elsewhere (2). Stories were scored by J. W. Atkinson whose scoring reliability is above .90 (2). The affiliation motive is best defined in terms of the content of imagery scored in the imaginative stories the essence of which is concern over establishing, maintaining, or restoring positive affective relations with other persons.

Experimental conditions. Approximately a month after the motive measures were obtained, the groups were introduced to the perceptual task. One whole class was introduced to this task with no immediately preceding attempt to arouse the affiliation motive. A second class participated in a sociometric rating procedure immediately before being introduced to the perceptual task. In this procedure, while all Ss were seated together around a large room, each S was required to place before him a card bearing a code number. Then he was required to make a series of ratings of his classmates of more than a month's acquaintanceship. First they were asked to rank-order a list of traits according to the degree to which possession of the trait would make a person likable. Then each was asked to pick from the list the two traits which best described each other person in the room. Finally, each was asked to write down the code numbers of the three persons whom he would most like to have as close personal friends. This task took about 20 minutes. In another year, half of the class was run under the neutral condition and the other half under a similar aroused condition. In an effort to keep the sociometric rating task and the perceptual task close together in time without establishing an apparent relationship between them, the perceptual task was introduced immediately without the usual break in the two-hour laboratory period. The excuse given was that time was running out and that there was one other task to be completed within the regular laboratory period. It was implied that the two tasks were unrelated.

The perceptual task. The Ss were randomly seated before a screen placed in the front of the room with their backs to a slide projector with a camera-type exposure diaphragm located at the back of the room. They were told that pictures would be projected on the screen before them at a speed and illumination inadequate for clear recognition. They were told that a picture would appear in each of four quadrants (top, bottom, left, and right), and though the image they received was very fuzzy and inadequate for recognition, they would be asked to designate in which of the four quadrants the picture *stood out the most* or was *most clear* to them. They were asked to write down a symbol for the appropriate quadrant after each of 80 exposure trials.

The psychophysical method, adopted from Blackwell (3), is suitable for measuring discrimination without conscious awareness. Blackwell has reported detection thresholds of a highly reliable character obtained with this method when the S was quite positive that no stimulus difference actually existed in the situation. In the present experiment, the exposure interval and level of screen illumination were both deliberately set at very low values on an empirical basis in pretest trials with other Ss, so that nothing more than a faint blur could be seen by the Ss. Many Ss refused to believe that there was any picture on

the screen during the exposure interval. Subsequent questioning of the Ss failed to reveal any who could give an adequate description of the composition of the exposure stimuli. For the most part, it can then be assumed that most Ss were performing their task at a perceptual level somewhat below the recognition threshold.

The stimuli. Each slide had four figures somewhat removed from the fixation point and equal distances up, down, left, and right from it. On every slide, one figure was that of a single face or several faces, and the other three figures were presumably neutral in the sense that they were not stimuli related to n Affiliation. Most of the neutral stimuli were pictures of home furnishings, e.g., a lamp, a plate about the size of a face, and in as far as possible having perceptual properties roughly equivalent to the pictures of faces.

There was a total of five face pictures and 15 neutral pictures. Each face-picture appeared on four slides, once in each available position, and neutral stimuli appeared at random in the other three positions. This arrangement produced five sets of four slides each, for a total of 20 slides. The 20 slides were presented in a predetermined random order, then in a second and independently selected random order for a total of 40 trials. These two orders were repeated to make a total of 80 trials. Each exposure was preceded by a verbal signal, "Ready now," in order to permit the Ss to focus on the fixation point in the middle of the screen.

Treatment of the data. The conditions under which the perceptual task was executed varied somewhat from group to group because the experiment was carried out in a classroom in which adequate control of ambient illumination could not be exercised. For this reason each experimental group was treated separately. For each S, the number of times he selected the face or faces in 80 trials was determined. Z scores were obtained for each of these distributions and it was then possible to place together groups which experienced slightly different conditions.

The n-Affiliation score in different groups was obtained from slightly different sets of pictures and from either four or six pictures. For this reason the n-Affiliation scores for each condition-homogeneous group was broken into High (above the median), and Low (below the median) n-Affiliation groups, and also into High, Middle, and Low (thirds) n-Affiliation groups.

RESULTS

The primary result may be seen in Table 1. Those Ss who are high in n Affiliation tend to choose faces more frequently than those Ss who are low in n Affiliation. The difference is significant at the .02 level. Although not reported here in tabular form, a similar difference at approximately the same significance level appears when the group is divided into High, Middle, and Low on n Affiliation.

Our treatment of the data permits us to

TABLE 1

N AFFILIATION AND THE PERCEPTION OF FACES
(Frequency that quadrant containing face(s) was
reported most salient in terms of Z scores from
neutral and motive-aroused
conditions combined)

	n-Affiliation Score	
	High	Low
Mean z score	.20	-.23
σ	1.01	.94
N	50	43
σ diff.	.205	
t	2.09	
p	.02 (one-tailed test)	

TABLE 2

PERCEPTION OF FACES AS A FUNCTION OF STRENGTH
OF AFFILIATION MOTIVE AND
EXPERIMENTAL CONDITION

(Frequency that quadrant containing face(s) was
reported most salient, in terms of Z scores from
neutral and motive-aroused conditions
presented separately)

n-Affiliation Score	Perceptual Condition			
	Neutral		Aroused	
	N	Mean z score	N	Mean z score
High	16	.14	17	.21
Middle	11	.15	20	-.01
Low	13	-.11	16	-.37
σ diff. (High vs. Low)		.38*		.35*
t		.66		1.67
p		n.s.		.05 (one-tailed)

* Derived from estimate of variance within groups = 1.0267
with $df = 87$.

compare all Ss who performed the perceptual task under neutral conditions with all those who performed it under aroused conditions even though each total group is composed of one semester's whole class and a half of another. The results of this analysis may be seen in Table 2. Here the original finding is seen again in the aroused condition but to a smaller and not significant extent in the neutral condition. This difference between aroused and neutral conditions, although large to the eye, .58 as compared to .25, is not statistically significant (σ diff. = .53). Furthermore, when this comparison is made with only a High-Low break of the n-Affiliation distribution the apparent difference is even smaller.

DISCUSSION

The clear relationship found between n-Affiliation scores and the tendency to select faces rather than neutral stimuli in a perceptual task seems to support a relationship between motivation and perceptual selection of motive-relevant stimuli. Since the measure of perceptual sensitivity in this case did not involve words, it seems reasonable to reject as an explanation any alternative based on verbal habits as suggested by Farber (4) or one based on frequency of word usage such as that offered by Solomon and Howes (8). Indirectly, this study lends support to the original interpretation in terms of motivation and perceptual selection of some of the studies that have involved the use of words. For example, the McClelland and Liberman study (6) relating n Achievement to recognition threshold for achievement-related words could properly be explained on the basis of verbal habits when considered alone. However, when they are viewed in comparison with the present study where words were not a factor in perceptual selection, it seems doubtful if an appeal to verbal habits constitutes a sufficient explanation.

The clear-cut primary results seem also to add to the evidence for the validity of the n Affiliation scoring system, since the scoring system yields statistically significant results in the expected direction on another variable.

The findings with respect to neutral and aroused conditions were not significantly different from chance. However, they are in the expected direction and the magnitude of the difference is approximately what was expected in comparison to the primary result. That is, there is some relationship in the neutral condition and more in the aroused condition. Since this is the second study in which the identical finding has been obtained (see also Moulton *et al.* [1]), it seems likely that the theoretical expectation (1 p. 90) will eventually be substantiated on the basis of frequency of appearance of the finding or in a study in which there is a very large N.

SUMMARY

Thematic apperception measures of n Affiliation were obtained from 93 male members of a class in elementary experimental psychology. Subsequently they were asked to say which of four figures which were flashed

on a screen in the four quadrants from the fixation point of up, down, left, and right, stood out the most or was clearest. The exposure speed and level of illumination was such that the stimuli were below the recognition threshold. On each trial one of the four stimuli was a face or faces and the others were similar but affiliation-neutral stimuli. It was found that those high in *n* Affiliation selected faces significantly more frequently than *Ss* low in *n* Affiliation. The expectation that arousal of the affiliation motive just prior to participation in the perceptual task would produce greater differences between High and Low *n* Affiliation *Ss* was confirmed but not to a statistically significant degree. This study is interpreted as supporting the predicted relationship between motivation and the perceptual selection of motive-relevant stimuli, and as lending indirect support to this interpretation of similar findings using word-recognition thresholds.

REFERENCES

1. ATKINSON, J. W. Explorations using imaginative thought to assess strength of human motives.

- In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln, Nebr.: Univer. of Nebraska Press, 1954.
2. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.
3. BLACKWELL, H. R. Psychophysical thresholds: Experimental studies of methods of measurement. *Engng Res. Inst. Bull.*, No. 36. Ann Arbor: Univer. of Michigan Press, 1953. Pp. 227.
4. FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955 **52**, 311-327.
5. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L., *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
6. MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236-251.
7. POSTMAN, L. J. The experimental analysis of motivational factors in perception. In M. R. Jones (Ed.), *Current theory and research in motivation*. Lincoln, Nebr.: Univer. of Nebraska Press, 1953.
8. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.

Received August 16, 1955.

CONTRASTING PATTERNS OF FANTASY AND MOTILITY IN IRISH AND ITALIAN SCHIZOPHRENICS

JEROME L. SINGER

*Franklin D. Roosevelt V. A. Hospital,
Montrose, N. Y.*

AND

MARVIN K. OPLER^{1, 2}

*Department of Psychiatry, Cornell University
Medical College and New York Hospital
(Payne Whitney Psychiatric Clinic)*

THE occurrence of differential patterns and manifestations of psychiatric illness in various cultures has long been documented by anthropologists and sociologists (2, 3, 10, 16, 17). There have been, however, relatively few studies employing psychological testing techniques to evaluate evidence of variations in personality characteristics of mental patients within American subcultural groups. The present study, part of a broader comparison of the psychopathology of two such subcultural groups, represents an initial effort to test certain hypotheses within a sample of schizophrenics stemming from Italian and Irish backgrounds. Underlying this study is the assumption that certain factors in early family structures and subsequent cultural conditionings have led to different modes of energy distribution and interpersonal orientation in Americans of Irish and Italian background and that these divergencies persist even when serious emotional disturbance is in evidence.

The specific focus of the present investigation is upon an aspect of personality tentatively designated as *fantasy-motility*. A considerable

body of research has suggested that motor and fantasy behavior may be viewed to some extent as functioning vicariously, the inhibition of motor activity leading to heightened fantasy or motion perception or the characteristic resort to fantasy being linked in many individuals with tendencies or capacities for motor inhibition (18, 21). The theoretical significance of this dimension stems from its relationship to a number of important theories of ego-development and the genesis and function of thought in human behavior (4, 8, 21). Much of the empirical work in this field has resulted from studies with the Rorschach inkblots and there is evidence that normals and schizophrenics who see relatively more Human Movement percepts (*M*) on the inkblots tend to show less overt movement, on the one hand, and to be more disposed to imaginative thought, on the other (4, 21, 22). To the extent that Rorschach's *M* response and its correlates may be taken as measures of basic ego functions involving the capacity to defer gratification and to resort to thought as "experimental action" (8), their employment in studies of subjects with differing cultural backgrounds has practical as well as theoretical implications.

There are suggestions in the literature that early identifications with significant family figures may be basic to the development of delaying capacities or fantasy (9, 12, 17, 21, 23). There is also evidence which points to differences in cognitive attitudes respecting time that seem to stem from differing family constellations in middle- and lower-class normals and schizophrenics (11, 13, 15, 20). These socioeconomic class differences cut across the normal-pathological dichotomy and it seems clear that, at least in response to projective test situations, the adult schizophrenic's pattern of "time perspective" and parental attitudes reflects in part cultural influence on family constellations as well as specific pathogenic childhood identifications. It is reasonable, therefore, to anticipate persisting personality differences

¹ Visiting Professor of Anthropology, Department of Psychiatry (Social Psychiatry), Cornell Medical College and Payne Whitney Clinic.

² This investigation represented a collaboration between personnel from the Veterans Administration Hospital and the Department of Social Psychiatry of Cornell University Medical College. It was in part supported by a research grant from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. Gratitude is also expressed for support from the Milbank Memorial Fund, the Grant Foundation, the Rockefeller Brothers Fund, and the Corporation Trust Company. Dr. Harvey J. Tompkins, as Chief of the Neuropsychiatric Section in the Veterans Administration, facilitated this collaboration. The authors are indebted to Dr. Harold Wilensky who obtained the behavioral ratings and to various staff members and trainees of the VA Hospital for assistance in obtaining the data. A separate report (18) deals with psychopathology and dynamic material derived from case histories, other tests, and intensive anthropological interviews.

among schizophrenic adults who come from cultural backgrounds in which early family structures differ and in which adult values and customs continue to stress divergent patterns of inhibition or deferment of gratification. Within the crude nosology of severe emotional disturbance, one may still expect to observe characteristic patterns of fantasy and motility that reflect cultural origins and that incorporate the gross symptomatology of schizophrenia, the hallucinations, delusions, the relative flattening of affect, and thought disorders within the context of these culturally-conditioned ego functions.

The selection of Irish- and Italian-Americans as subjects for the study was based on accumulating anthropological and sociological indications of marked differences in their cultural patterns, family constellation, and male social and sexual role positions (1, 2, 3, 5, 16, 17). The clinical experience of the authors based on intensive study with individuals from these cultural groups further amplified the general picture as did considerable evidence from the existing literature. In the Irish family an oft-observed central and controlling maternal role contrasts with that of the Italian mother who, despite warmth and protectiveness, often represents a reflection of greater paternal authority. Sexuality in the two traditions differs as do attitudes towards their common religion. In the Italian family sexuality is often emphasized as part of the expressive and emotional life and is accepted as an assertion of maleness. Among the Irish, known for their mild and protracted courtships, delayed marriages, and celibate emphases, sexuality is clearly subordinated to procreation and, apart from this framework or from the marital setting, is more likely to be regarded as guilt-producing and sinful. In brief, the Irish family constellation with the relatively powerful mother-figure and female siblings tends to emphasize inhibition and delay of gratification with ambivalence towards the female and difficulty in establishing an identification with a father-figure who is psychologically neutral and frequently absent from the home.

In the Italian family, fairly common practice permits use of the normative cultural pattern for an expressive acting out of feelings. Thus, the emphasis is on direct expression of emotions resulting from conflicts with the relatively powerful father and older male siblings. Such

emotions may tend in their pathological extremes to be built more readily into hostile patterns of reaction, often self-destructive as well as destructive of the male parental image. A previous anthropological survey by Opler of milder forms of mental illness among Irish and Italians in another hospital had suggested that the Irish male schizophrenic may be seen as beset with fear and guilt, controlling deeply repressed hostility towards female images chiefly through repression and fantasy. The Italian male model for schizophrenia seemed, on the contrary, given to more destructive and overtly hostile impulses, aimed chiefly at the male parental image but built up to uncontrollable dimensions in accordance with a cultural value of emotional expression at all costs.

In proceeding toward an operational formulation of hypotheses derived from the cultural differences between the Irish and Italian, it was decided to employ a battery of evaluative procedures with fairly well known properties. Rorschach's Human Movement response (*M*) seemed a logical variable for study as well as the Thematic Apperception Test Transcendence Index (22), both measures of fantasy disposition. A time-estimation task involving delaying ability for accuracy as well as a slow-writing task were included as measures of inhibition. Similarly, the Porteus Mazes, involving as they do a combination of motor restraint and foresight for successful performance, were included in the battery. The resort to fantasy and greater obsessional tendencies of the Irish might be expected to lead to more ready *admission of disturbance* following failure in a frustration-tolerance situation. Finally, in order to obtain something approximating a longitudinal section of behavior in the motor area, ratings of ward behavior with respect to cooperativeness and aggression or assaultiveness were employed (14).

HYPOTHESES

1. Irish schizophrenic patients differ from Italian schizophrenics by showing greater evidence of fantasy-tendency. Specifically, they show more Human Movement responses on the Rorschach inkblots, lower threshold for Human Movement on Barron's inkblots, and higher scores in Transcendence on Thematic Apperception Test stories.

2. Irish patients show greater evidence of inhibition, motor inhibition, and restraint on

tasks calling for delaying capacity. They require longer times in a slow-writing motor inhibition task, score higher Test Quotients on the Porteus Mazes, and achieve greater accuracy in Time Estimation in a situation involving delay in response.

3. Irish patients more readily admit to disturbance following failure in a frustration-tolerance test.

4. In ratings by personnel of their ward behavior, Irish patients are less aggressive, less overtly assaultive, and more cooperative in ward routines than are Italian schizophrenics.

METHOD

Subjects

A total of 60 male veteran patients diagnosed as schizophrenic were chosen from the wards of the hospital so as to constitute two samples of thirty patients each from Irish-American and Italian-American cultural backgrounds. Criteria for inclusion of patients involved (a) use only of clearly diagnosed schizophrenics without known organic brain damage, (b) first to third generation from both ethnic groups with residence in the New York City metropolitan area, (c) ages 18-45 inclusive, (d) patients sufficiently in contact to undergo the test battery and anthropological interview. Table 1 presents pertinent data comparing the two samples for age, education, date of first hospitalization, and Wechsler-Bellevue IQ. An evaluation of socioeconomic status was also carried out, using as indices parents' education and occupation, income, and housing as well as Ss' standing on these variables. Socioeconomic status and religion (all Ss were professed Roman Catholics) did not differ for the two samples. Within the Irish group, all families originated in the poorer Southwest counties while the Italian Ss with two exceptions were of South Italian or Sicilian origin. Generation level was approximately equated in each group, thus limiting differential effects of acculturation, a factor of considerable significance for motor patterns in Efron's work (7).

Procedure

All Ss received an assessment battery administered in standard fashion. The following is a brief description of the procedures and the measures employed to test experimental hypotheses:

Rorschach Inkblots. All Rorschachs were scored by Klopfer's method. The primary variable derived from the Rorschach was the Human Movement response (*M*). Only frequency of occurrence was considered in this study. Statistical analysis of other Rorschach data was limited to those variables in which a sufficient range of scores emerged.

Barron's Movement-Threshold Inkblots. These blots were included in order to evaluate their suitability as a measure of Rorschach *M* more easily obtained and more suitable for quantification. Barron's inkblot

series (4) consists of a series of black-and-white inkblots arranged in a sequence which elicits increasing numbers of *M* responses over the range of 30 cards. The score employed was the card number on which the initial Human Movement response was offered by S.

Thematic Apperception Test. Four cards, 1, 6, 7, and 12, of the Murray series were administered. Weisskopf's Transcendence Index as employed in an earlier study (22) was applied to the content of the stories. This index, a measure of the extent to which S's story introduces characters, activities, and emotions not actually present on the stimulus card, may be thought of as a reasonable operational definition of fantasy disposition. The actual score employed was the total number of transcendent items over the four cards.

Porteus Mazes. This instrument was administered in standard fashion. Test Quotients were computed in accordance with norms available from Porteus (19). Qualitative notes on mode of approach were kept.

Time Estimation. As a fairly direct measure of delaying capacity, the Ss were asked to indicate when they thought a given time interval had elapsed. Three times were used, 15, 30, and 60 seconds. The *E* started a stop watch and recorded the actual time passed when S signaled that he thought the stated interval had passed. The score was the sum of the three estimations subtracted from absolute times. Since almost all Ss indicated that the time had passed before it actually had, accuracy and delaying ability were practically identical.

Motor Inhibition. This task, adapted from Downey's Will-Temperament scale, required Ss to write a simple phrase as slowly as possible without stopping the motion of their pencils. Time in seconds taken to write the phrase was the basic score.

Admission-Denial of Frustration. This variable was derived from response to a simple experimentally induced failure experience devised by Wilensky (24). Following a period of nine successive failures on digit series, Ss almost invariably showed impairment in retention of digit series previously within their scope. Despite this apparent evidence of disturbed performance following failure, many Ss denied conscious distress upon direct inquiry. In this study each S was asked if he had been disturbed by his failures. Responses were categorized as *admissions* or *denials* of disturbance following failure. Some cases were lost in this technique because of bizarre verbalization or refusals to comment.

Behavior Ratings. To obtain some evaluation of a longitudinal nature, two groups of items from the Multidimensional Rating Scales for Ward Behavior of Psychiatric Patients (14) were employed. These items, grouped into scales for Aggression and Cooperation, were rated by a staff clinical psychologist during interviews with ward personnel, aides, nurses, and physicians, all of whom had observed the patients for approximately six months on their wards. The emphasis in these scales is on actual incidents and overt behavior rather than on interpretations of underlying trends. Items on the Aggression Scale included instances of overt destructive behavior, assaultiveness, and verbal outbursts, while items on the Cooperation Scale involved such factors as willingness to follow ward thera-

TABLE 1

MEANS OF IRISH AND ITALIAN SCHIZOPHRENICS ON PERTINENT BACKGROUND FACTORS, RORSCHACH DETERMINANTS, AND EXPERIMENTAL VARIABLES

Variables	Irish (N = 30)	Italian (N = 30)
Background Factors		
Age	32.0	30.5
Educational Grade Level	10.5	10.9
Year of First Hospitalization	1949.8	1949.5
Wechsler-Bellevue IQ	108.4	105.5
Rorschach Variables		
Total Responses (<i>R</i>)	21.1	20.7
Human Movement (<i>M</i>)	2.2	1.3
Animal Movement (<i>FM</i>)	3.4	2.8
Form-Color (<i>FC</i>)	0.8	0.7
Color-Form (<i>CF</i>)	1.5	1.4
Weighted Color Sum (<i>Sum C</i>)	2.5	2.6
Experimental Variables		
Movement-Threshold Inkblots	14.3	20.8
TAT Transcendence Index	19.3	14.3
Porteus Mazes TQ	87.2	86.7
Time Estimation (dev. in secs. from absolute times)	38.8	44.0
Motor Inhibition Time (secs.)	147.0	106.5
Admit Frustration	67%	42%
Aggressive Ward Behavior Rating	1.60	2.30
Cooperation Behavior Rating	1.87	0.67

peutic routines, assisting in ward chores, and accepting authority of ward personnel. As employed here, high scores on these scales imply considerable aggressive behavior and minimal cooperation.

RESULTS

Table 1 presents a comparison of Irish and Italian group means for background variables, Rorschach determinants, and experimental variables. The groups do not differ significantly in age, years of education, date of first hospitalization (an estimate of chronicity), Wechsler-Bellevue IQ, or in the representation from various socioeconomic levels. Evaluation of significance levels was carried out by dichotomizing Irish and Italian distributions for all variables at the medians and employing chi square to evaluate the association of cultural background with high or low scores on a given measure. Following Edwards (6), r_{ϕ} was calculated between Irish-Italian on each measure as a rough estimate of extent of association. A positive correlation indicates higher scores for Irish Ss on the particular variable. Results of these analyses are presented in Table 2.

Inspection of the tables suggests consistent support for the hypotheses. The Irish and

TABLE 2

VALUES OF χ^2 AND r_{ϕ} FOR DICHOTOMIZED DISTRIBUTIONS OF IRISH AND ITALIAN GROUPS ON PERTINENT BACKGROUND, RORSCHACH, AND EXPERIMENTAL VARIABLES

Variable	χ^2	r_{ϕ}
Age	0.67	+.03
Education	0.13	.00
Year of first Hospitalization	0.0	.00
Wechsler-Bellevue IQ	1.80	+.17
Rorschach Responses	.13	.00
Rorschach <i>M</i>	8.4**	+.59
Movement-threshold	11.3**	-.68
TAT Transcendence Index	6.7**	+.53
Porteus Mazes	0.67	-.03
Time Estimation	4.0*	-.40
Motor Inhibition Time	5.0*	+.45
Admission-Denial of Frustration	3.1*	+.38
Aggression Behavior Rating	6.1**	-.51
Cooperation Behavior Rating	3.4*	+.38

* Significant at $p/2 < .05 > .01$

** Significant at $p/2 < .01$

Italian Ss do not differ in frequency of the Rorschach determinants except for the *M* responses, which are definitely more numerous among the Irish. The Irish patients, as hypothesized, show lower thresholds for Human Movement responses in Barron's inkblots, greater transcendence in their TAT stories, lower deviation from absolute times in Time Estimation, longer inhibition times in the slow-writing task, greater tendency to admit distress following failure, and, in their ward behavior, less aggression and greater cooperativeness. Only the Porteus Maze scores reveal no differences and fail to support the relevant hypothesis. A qualitative difference in performance on the mazes was evident, however. While both groups obtained similar scores, the Irish, according to qualitative notes, were far more often reported to be deliberate and cautious in performance, whereas the Italian Ss were more often described as impulsive and rapid. Why this difference did not affect ultimate outcome in maze score could not be ascertained.

In connection with the ward ratings, it should be noted that ward personnel who provided the data and the raters were ignorant of the purposes of the experiment. Similarly, the difference between the groups in ward cooperation could not account for test score differences since, in the actual testing situation, the Italian Ss proved to be somewhat more genial and amenable to examination.

DISCUSSION

There seems little question that the test performance of these two groups reflects clearly a difference in fantasy and motility that is strikingly in accord with anticipations based on anthropological evaluations of cultural frameworks and family constellations. When it is kept in mind that both the Italian and Irish patients are American citizens, most of them native born, that they have participated in the same urban culture and have also undergone the standardization of military service, then the persistence of differences rooted in subcultural group participation is highly noteworthy. When it is further recognized that both groups of patients were grossly psychotic, the emergence of the differential fantasy-motility pattern is indeed a remarkable demonstration of the hypothesized group differences. The patients in this study all showed many symptoms of schizophrenia, although such symptoms as "flattening of affect" were less apparent in the Italian and "catatonic excitements" were not characteristic of the Irish. At the same time, despite the obvious fact that thinking disorders and hallucinations were present in most cases, wide variations in personality structure emerged. As James S. Plant, Fromm-Reichman, and Horney have suggested, however, these data and other materials from the same larger study (18) suggest that schizophrenias and other personality disorders highlight culturally engendered conflicts. This view poses serious questions about the validity and usefulness of current psychiatric nomenclature (16). Should study with further samples confirm the findings of this report, extensive familiarity with subcultural patterns may be increasingly essential for effective psychiatric diagnosis and personality evaluation. The need for further work using test batteries like the one reported here with normal Ss from various subcultural groups is obvious.

Pending further exploration it seems reasonable to conclude that schizophrenics stemming from Irish and Italian cultural backgrounds do show markedly different patterns of fantasy and motility. The results seem best interpreted as suggesting the possibility that because of the limitation on variability in family constellation imposed by cultural background, as well as the recurrent influences of cultural values

during development, characteristic differences in modal personalities (9) from different subcultural groups persist through adulthood and are in evidence even during a period of severe emotional disturbance. To the extent that such modal personality patterns may be derived from differing family constellations and cultural values, an extensive exploration of fantasy and motor activity as well as of other relevant personality dimensions by means of joint psychological testing and anthropological survey with many other subcultural groups appears to be desirable. Such explorations might go beyond traditional cultural stereotypes toward an understanding of variables underlying the relationships between patterns in family life and cultural standards as they interact to mold the adult personality in its normal and pathological manifestations.

SUMMARY

This study represented a portion of a larger investigation involving a comparison of the psychological test performance and psychopathological manifestations of schizophrenic subjects from different American subcultural groups. Based on anthropological evaluation of cultural patterns and child-rearing practices in families of Irish and Italian ethnic background, certain hypotheses concerning differences in fantasy and motor activity were developed. The Irish were expected to prove more given to imaginative behavior and motor control when compared with Italian Ss. When comparable samples of Irish and Italian schizophrenic males were studied by psychological tests and ward behavior ratings, these hypotheses were essentially confirmed. In the test results, no differences emerged in Rorschach variables except for the Human Movement responses, more of which were produced by the Irish. The Irish showed lower threshold for perception of Human Movement on Barron's inkblots, greater transcendence of stimulus content in their TAT stories, longer motor inhibition times on a slow-writing task, longer delay in time estimation, greater tendency to admit distress following frustration. Ratings of ward behavior indicated the Irish as more cooperative in therapeutic routines and less overtly aggressive than the Italian patients. No differences in Porteus Maze Test Quotients emerged, although a qualitative tendency for

the Irish to work more slowly and hesitantly compared with the impulsivity and haste of the Italian Ss was clearly apparent. In general, the results appear to support the hypothesis of persistent differences in motor and fantasy activity between patients from the two ethnic groups. Implications for review of psychiatric nosology, personality research, and interdisciplinary collaboration are manifold.

REFERENCES

1. ARENSBERG, C., & KIMBALL, S. T. *Family and community in Ireland*. Cambridge: Harvard University Press, 1948.
2. BALES, R. F. Cultural differences in rates of alcoholism. *Quart. J. Stud. Alcohol*, 1946, **6**, 480-493.
3. BARRABEE, P., & VON MERING, O. Ethnic variations in mental stress in families with psychotic children. *Soc. Problems*, 1953, **1**, 48-53.
4. BARRON, F. Threshold for the perception of Human Movement in inkblots. *J. consult. Psychol.*, 1955, **19**, 33-38.
5. CHILD, I. L. *Italian or American?* New Haven: Yale University Press, 1943.
6. EDWARDS, A. *Statistical analysis for students in psychology and education*. New York: Rinehart, 1946.
7. EFRON, D. *Gesture and environment*. New York: Kings Crown Press, 1941.
8. FREUD, S. Formulations regarding the two principles in mental functioning. *Collected Papers*, Vol. IV. London: Hogarth, 1946.
9. INKELES, A., & LEVINSON, D. J. National character: the study of modal personality and socio-cultural systems. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 977-1020.
10. KLUCKHOHN, C. Culture and behavior. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 921-976.
11. KOHN, M. L., & CLAUSEN, J. A. Parental authority behavior and schizophrenia. *Amer. J. Orthopsychiat.*, in press.
12. LAIR, W. S. *The psychoanalytic theory of identification*. Unpublished doctor's dissertation, Harvard University, 1949.
13. LE SHAN, L. L. Time orientation and social class. *J. abnorm. soc. Psychol.*, 1952, **47**, 589-592.
14. LORR, M. Multidimensional scale for rating psychiatric patients. Hospital form. *VA tech. Bull.*, No. 10-507, Nov. 16, 1953.
15. MITCHELL, H. E. Social class and race as factors affecting the role of the family in Thematic Apperception Test stories of males. Unpublished doctor's dissertation, University of Pennsylvania, 1950.
16. OPLER, M. K. Cultural anthropology and social psychiatry. *Amer. J. Psychiat.*, in press.
17. OPLER, M. K. Cultural perspectives in mental health research. *Amer. J. Orthopsychiat.*, 1955, **25**, 51-59.
18. OPLER, M. K., & SINGER, J. L. Ethnic behavior and psychopathology: Italian and Irish. *Int. J. soc. Psychiat.*, in press.
19. PORTEUS, S. D. *The Porteus Maze test and intelligence*. Palo Alto: Pacific Books, 1950.
20. SINGER, J. L. Projected familial attitudes as a function of socioeconomic status and psychopathology. *J. consult. Psychol.*, 1954, **18**, 99-104.
21. SINGER, J. L. Delayed gratification and ego-development: implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19**, 259-266.
22. SINGER, J. L., & HERMAN, J. Motor and fantasy correlates of Rorschach Human Movement responses. *J. consult. Psychol.*, 1954, **18**, 325-331.
23. SINGER, J. L., & SUGARMAN, D. A note on some projected familial attitudes associated with Rorschach movement responses. *J. consult. Psychol.*, 1955, **19**, 117-119.
24. WILENSKY, H. The performance of schizophrenic and normal individuals following frustration. *Psychol. Monogr.*, 1952, **66**, No. 12 (Whole No. 344).

Received August 16, 1955.

THE EFFECT OF GOAL VALUE UPON EXPECTANCY¹

LEONARD WORELL

Reed College

BEHAVIORAL expectancies do not always appear to be governed by "realistic" factors in a situation, but seem also to be a function of wishes, fears, and doubts. Observers disagree, however, about the manner in which differentially valued goals affect expectancies. Some believe, for example, that expectancies are lowered, in the face of highly valued goals, as evident in the student before an important examination, while others maintain that they are raised, as in gambling situations. The present study attempts to examine some of the conditions under which expectancies may or may not be related to different levels of goal value.

This research was conceived within the framework of a social learning theory of personality developed by Rotter (11). Within this system predictions of behavior are mediated by two major constructs, expectancy and reinforcement value. Expectancy is defined as "the subjective probability held by the individual that a particular reinforcement will occur as a function of, or in relation to, a specific behavior in a given situation or situations" (11, p. 112). Expectancies are viewed as being determined by experience in particular situations and in related situations. Thus, a student taking a history course may have expectancies that studying will lead to a certain grade, based upon experiences in this history course, other history courses, other academic courses in general, and so on. The second construct, reinforcement value, is defined as "the degree of preference for any reinforcement to occur if the possibilities of their occurring were all equal" (11, p. 112). This value may be either

positive or negative. The terms goal value and reinforcement value are used interchangeably hereafter.

From a social learning point of view, there is no *necessary* relationship between expectancy and goal value. Under certain conditions a specific relationship may be anticipated, while under others the nature of the relationship is obscure. The purpose of this study is twofold, to pursue an empirical phenomenon of general interest and to contribute to the predictive power of Rotter's theory.

In examining the relationship of goal value to expectancy, previous investigators have consistently used mean expectancy scores derived from a series of trials. For this type of situation the nature of the relationship between the constructs cannot be logically deduced from Rotter's theory. From the standpoint of other systematic points of view and general observations, conflicting conceptions emerge. Thus, in Lewin's formulations (8), as the difficulty level of a situation increases the value of success increases, while the subjective probability of success attainment decreases. The analysis of individuals confronted with high value situations seems to bear this out, for low expectancies are generally elicited. The suggestion follows that confidence in attaining goals decreases as they become more important, possibly because the acquisition of high goals in the past has involved considerable effort and competence. On the other hand, it seems equally reasonable to suppose that expectancies increase in highly valued situations. Our society might be seen as providing a particularly striking example, in that both striving for high goals and displaying confidence in one's ability to achieve them are held as the ideal. In view of these alternatives, the exact nature of the relationship between goal value and expectancy over a series of experiences in the same situation must be a matter for empirical determination.

On the basis of Rotter's theory, at least two conditions should lead to the absence of any observable relationship between the constructs.

¹ Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, The Ohio State University. The writer wishes to express his appreciation to Professors Julian B. Rotter, Delos D. Wickens, Lauren G. Wispé and Robert J. Wherry for their helpful suggestions and advice. The writer is indebted to Russell Slater, director of the Child Study Division, and to the principals and teachers of the participating schools in the Columbus Public School System for their cooperation in making subjects available.

One of these hinges on an assumption of social learning theory that expectancies are a function of experiences in both a particular situation and in related situations. Expectancies in a relatively novel situation should thus be determined by experiences from related situations. Such related experiences may include the occurrence (or nonoccurrence) of past reinforcements and, possibly, the values of these reinforcements. With additional experience in a situation, expectancies should increasingly be determined by the actual occurrence of reinforcements in that situation (2, 4). Although the nature of the influence of different goal values upon expectancies cannot be specified in a novel situation, after experience with that situation expectancies should therefore be determined *only* by the occurrence and *not* by the value of the reinforcements.

The second condition in which differential goal values should produce little or no effect upon expectancies deals with the introduction of a penalty for inaccuracy in stating expectancies. Relatively widespread use has been made of penalties in level-of-aspiration experiments. According to social learning theory, a penalty may be viewed as a reinforcement value associated with accuracy. This additional reinforcement value (often interpreted as a negative reinforcement by the *S*) should counteract the tendency to make cautious or wishful expectancy statements engendered by high or low goal values. If the severity of the penalty is identical for different goal value conditions, therefore, expectancies should be uniform in relation to these goal values. In addition, expectancies should be more realistic or performance-oriented than in nonpenalty conditions.

In summary, three aspects of the relationship between goal values and expectancies are under investigation. The question is first asked whether (a) differential goal values affect expectancies over a series of experiences in the same situation. The remaining two aspects involve attempts to support extensions of Rotter's social learning theory. Thus, (b) the relationship, if any, observed in *a* is hypothesized to be a function of the amount of experience in the situation. Relatively novel situations may elicit a relationship between expectancies and differential goal values, but experience in these situations should produce

uniform expectancy statements. Finally, (c) it is hypothesized that the introduction of a reinforcement value for accuracy leads to more realistic expectancies, on the one hand, and expectancies that are unaffected by the value of an event, on the other.

METHOD

Selection of Tasks

A number of requirements had to be met in the selection of tasks. First, the tasks had to be relatively novel, so that expectancies would not be established pre-experimentally. Second, they had to permit *E* to control performance, in order that the reinforcements would be uniform for all *Ss* under all value conditions. Finally, they needed face validity for the goal area involved, i.e., athletic skills. From a pilot study, three tasks were chosen, each one representing a different goal-value level. These were (a) the Rotter level-of-aspiration board (10) modified by a cover placed over the numbers so that *S* was unable to see his performance; (b) a rotary pursuit test (score measured by amount of time that the stylus is kept on revolving disc); and (c) a set of six blocks for a tapping test. The results of the pilot study indicated that the tasks were perceived as having differential goal values. Through a ranking procedure, it was found that the level-of-aspiration board and the rotary pursuit task were more highly valued than the tapping task. These initial preferences were coupled with instructions (see below) that assigned more or less value to each task.

Subjects

A total of 102 boys, ranging in age from nine to eleven, was selected from the fifth and sixth grades of the local public school system. Eight *Ss* who were unable to comply with the instructions were discarded.

Administration of Tasks

Each *S* was seen individually. He was seated at a table on which the three tasks were arrayed. The *S* was given a description of each task in terms of its relative merit in providing information about a person's athletic abilities. Three levels of goal value were employed, designated as high, medium, and low. For the high reinforcement-value task (level-of-aspiration board), *S* was told that this was the most important task, that it provided the most information about skills required in all sports, e.g., coordination, timing, and balance, and that this task was of greatest concern to *E*. The medium value task (rotary pursuit) was said to be not quite as useful and to provide less information about a person's abilities. Finally, the low value task (tapping) was described as much inferior to the others and of almost no interest to *E*. When tapping was followed by one of the other tasks, *S* was also instructed to "get this one over quickly, so that we can get on to the others."

A level-of-aspiration procedure was used in obtaining expectancies. The *S* was informed of the range of possible scores (0 to 50) and was asked to provide an estimate of his subsequent performance for each trial.

TABLE 1
MEAN *D* SCORES AND STANDARD DEVIATIONS FOR ALL GROUPS

Goal Value	Nonpenalty Group						Penalty Group	
	Six Trials		1st Trial		6th Trial		Six Trials	
	M	SD	M	SD	M	SD	M	SD
High	5.5	16.2	.9	4.6	.3	2.8	-3.7	14.6
Medium	11.6	17.1	3.8	4.9	.7	2.7	-.5	15.7
Low	10.6	14.6	3.0	4.5	1.0	3.0	2.8	14.6

TABLE 2
MEAN DIFFERENCES AT EACH COMBINATION OF REINFORCEMENT VALUE

Group	<i>N</i>	Reinforcement Value Conditions								
		High and Medium			High and Low			Medium and Low		
		\bar{D}	<i>t</i>	<i>p</i>	\bar{D}	<i>t</i>	<i>p</i>	\bar{D}	<i>t</i>	<i>p</i>
Nonpenalty	71	6.1	3.82	<.001	5.1	3.53	<.001	1.0	.54	NS
	71	2.9	4.10	<.001	2.1	2.91	<.01	.8	1.27	NS
1st Trial	71	.4	1.32	.20	.7	2.33	<.05	.3	.28	NS
6th Trial	71	.4	1.32	.20	.7	2.33	<.05	.3	.28	NS
Penalty	23	3.2	1.14	NS	6.5	3.01	<.01	3.3	1.92	.10

This was followed over the series of six trials. Tasks were rotated, so that *Ss* performed differentially valued tasks in varying orders. A score of either 26, 27, or 28 was given *S* on each task immediately after his first performance. Following this, a predetermined series of scores was supplied which provided the same mean increase in performance over the six trials for each of the three tasks.

Seventy-one *Ss* performed the tasks without a penalty for inaccuracy of estimates. Another 23 *Ss* performed the same tasks with a penalty being established for inaccuracy by instructing *S* that he would lose twice the number of points that his expectancy fell below his actual performance, and that he would obtain exactly what he expected if his actual score were at or above his expectancy estimate. Each *S* was given detailed examples and was not permitted to begin the tasks until he could calculate his score on an example without assistance.

RESULTS

Analysis of data was based on the stated expectancies to the three tasks of high, medium, and low goal value. A *D* score was employed, the difference between the expected score and the preceding performance. Tests of the hypotheses, therefore, are focused on the differences between mean *D* scores (Table 1) for each of the reinforcement-value conditions.

Effects of goal value. Bartlett's test of homogeneity of variance was applied before performing an analysis of variance and the assumption

of homogeneity was supported. The analysis of variance indicated that the difference attributable to reinforcement-value conditions is highly significant ($p < .001$). Two-tailed *t* tests for correlated means were used in comparing mean *D* scores. Two of the three comparisons are significant (Table 2); the mean *D* score at the high value is significantly lower than either the medium or low values.

Effects of experience. As six trials were used, differences between expectancies might be anticipated at the first trial, while none should appear at the sixth trial. Analyses of variance indicated that the reinforcement-value conditions continued to provide significant sources of variance at both trials ($p < .001$ at Trial 1, $p < .05$ at Trial 6). When mean *D* scores are compared at Trial 1 (Table 2), the differences between the high and medium values and the high and low values are highly significant. At the sixth trial, a significant difference is found only between the high and low value conditions.

Effects of penalty. Table 1 contains the mean *D* scores for groups with and without a penalty associated to expectancy statements. Inspection of this table confirms the hypothesis that a penalty leads to more realistic expectancies. "Realistic" is defined in terms of the mean *D* scores' proximity to zero, or stated in another way, the extent that expectancy statements approach previous performance. The differences between penalty and nonpenalty groups are significant in all cases, falling at less than the .02, .01, and .05 levels at the high, medium, and low values respectively.

Our interest here, however, lies also in the differences between mean *D* scores at the three value conditions of the penalty group alone. As before, the analysis of variance continued to show that a significant source of variance may be attributed to the reinforcement-value conditions ($p < .05$). In applying two-tailed *t* tests for correlated means (Table 2) to the differences between mean *D* scores, the only significant difference is found between high and low conditions, although the difference between medium and low conditions approaches significance. When the differences between means in the penalty group are examined more closely, it appears that the effect of a uniform penalty at all value conditions is to produce more pronounced separations between high, medium, and low goal values.

DISCUSSION

The most general statement that may be drawn from the results is that the value of a goal has an effect upon the level of stated expectancy. It was found that (a) the presence of high goal values leads to significantly lower expectancy estimates over a series of experiences in the same situation than do either medium or low value conditions; (b) the amount of experience, though appearing to dilute somewhat the effects of goal values upon expectancy, does not eliminate the tendency for high goals to be associated with lower expectancies in the six trials of the experiment; and (c) with the inclusion of a penalty (reinforcement value) for inaccuracy, contrary to what was anticipated, expectancies continued to be lower in relation to highly valued goals. A penalty did, however, lead to more realistic or performance-oriented expectancy estimates.

The influence of reinforcement value upon expectancy was not demonstrated at all levels of goal value. That expectancies were relatively lower for high values suggests that the frequently quoted expression "the good things are always hard to get" may well be applied here and that individuals appear to perceive high goals as difficult in the Lewinian sense. The failure to obtain significant differences between medium and low value conditions, however, requires explanation. A possible factor may have been the relative absence of perceived separation between the experimental goal-value conditions. The division between the three goal values was made in accordance with an expected theoretical and experimental gradient. It is not known, however, how distinct the goals were or how important the distinction seemed to the Ss. A completely satisfactory method for measuring precise distances between goal values has yet to be developed.

Goal values appeared to exert their strongest influence upon expectancies in relatively novel situations. This finding suggests that different initial expectancies are held in relation to varying goal levels from previous related situations. People seem to differentiate relatively novel situations not only in accordance with the possible occurrence (or nonoccurrence) of reinforcements, but also with regard to the values of the reinforcements. The behavior of the Ss in this study suggests that individuals may ordinarily respond in terms of two steps along

this continuum, e.g., high and low value. The presence of a penalty seemed to be a condition leading to more subtle discriminations between values.

Reference has already been made to the fact that the effects of goal value on expectancy did not disappear with experience. There was, however, a tendency for the influence of goal values to be diminished. This observation derives further support from a comparison of the mean changes that are found at each combination of reinforcement value between Trials 1 and 6. When the t test² is applied to these mean changes, the differences between trials for the high and medium ($t = 2.8, p < .005$) and the high and low ($t = 1.9, p < .03$) are highly significant, while that for the medium and low values approaches significance ($t = 1.6, p = .06$). This finding is interpreted to mean that differences between expectancies have been reduced at the sixth trial and that a more adequate test of the effects of experience would be made by extending the number of trials beyond six.

In securing a particular relationship between goal value and expectancy, this study may seem merely to have added another conflicting result to an array of studies already in disagreement (1, 3, 5, 7, 9). Representative of one kind of previous approach to the problem is a study by Marks (9), which sought to determine the desirability of outcome on the stated expectations of children in certain choice situations. Using what may be referred to as a quasi-gambling technique, she found that the value of an event exerted a strong effect upon stated expectations. Contrasted with this, Holt (5), employing what may be called a competitive level-of-aspiration technique, demonstrated no effect of "ego-involvement" on the stated expectancies of his college Ss. The present study adds to the picture of inconsistency.

A possible resolution may be achieved by differentiating the kinds of situations studied. At least two broad divisions of situations may be distinguished: *achievement* and *nonachievement*. In the former, performance is dependent upon ability or skill, and thus reflects upon an individual's competence. In the latter, which may include gambling, some play and simple

² Since these were anticipated outcomes of our hypothesis regarding effects of experience, one-tailed tests of the distribution of t were employed.

choice situations, no ability or skill is prominently involved, and an individual's competence is not challenged. One might anticipate differing results of the effects of goal value upon expectancy under these two conditions, since the achievement situation introduces additional goal values.

This distinction appears to assist in resolving disparities in the results of studies such as that of Marks and the present one. Her results strongly indicated that *increased* value led to *increased* expectations. According to the present analysis, her study would fall within the nonachievement category, since children were expressing what they *wanted* to happen; no ability or skill was implied. In this sense, her *Ss* were faced with a form of gambling, and culturally, one may find a greater tendency to take risks in these situations. In this investigation, performance was related to ability, so that an additional goal value was introduced. Expectancies held in relation to previous *achievement* situations would therefore be most relevant. The results of this experiment demonstrated that expectancies *decreased* with *increasing* value in relatively novel situations. Whether this proposal has any merit aside from providing *post hoc* explanation will be determined by future experimentation.

The findings were somewhat equivocal with respect to the penalty for inaccuracy of expectancy statements. Though expectancies were more accurate, they did not appear to be removed from the effects of goal values. Before conclusions can be drawn, it would seem necessary to investigate the problem of penalties more thoroughly. Different severities of penalties and the effectiveness of instructions in communicating penalties could well bear further examination.

SUMMARY

In Rotter's social learning theory of personality, predictions of behavior are made through the operation of two central constructions, expectancy and reinforcement value. The present study was designed to investigate the effect of different goal (reinforcement) values upon expectancy.

Three levels of goal value were employed in a study of fifth- and sixth-grade boys. Expectancies developed from related situations were controlled by having each *S* perform each of

three relatively novel tasks. In addition, the performance of *S* was controlled by giving prearranged sequences of scores. Thus, with both performance and experience from related situations controlled, the only factor varied was the value of each task.

The most consistent finding was that the value of an event has some effect upon stated expectancy. Further results of the study may be summarized as follows:

1. Expectancies observed over a series of experiences in the same situation were found to be significantly lower in highly valued situations as compared to situations of lower value.
2. Though the influence of different goal values upon expectancies appears to become more uniform with experience, expectancies continued to be significantly lower in high value conditions. The necessity of investigating more extensive amounts of experience than those considered here was suggested and discussed.
3. The association of a goal value for accuracy (penalty) to expectancy statements leads to more realistic expectancies, that is, expectancies tend to approach previous performance more closely. The penalty, however, did not appear to eliminate the effect of different goal values upon expectancy. Expectancies were found to be relatively lower to high goal values, as they had been in situations without a penalty.

Finally, in an attempt to provide a resolution of the seemingly contradictory results of previous studies in this area, the potential utility of a distinction between achievement and nonachievement situations was discussed.

REFERENCES

1. BAYTON, J. A. Interrelations between levels of aspiration, performance, and estimates of past performance. *J. exp. Psychol.*, 1943, **33**, 1-21.
2. DEAN, S. J. Sources of variance in individual statements of expectancy. Unpublished doctor's dissertation, Ohio State Univer., 1953.
3. FRANK, J. D. Some psychological determinants of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 285-293.
4. GOOD, R. A. The potentiality for change of an expectancy as a function of the amount of experience. Unpublished doctor's dissertation, Ohio State Univer., 1952.
5. HOLT, R. R. Effects of ego-involvement on levels of aspiration. *Psychiatry*, 1945, **8**, 299-317.
6. IRWIN, F. W. Stated expectations as functions of probability and desirability of outcomes. *J. Pers.*, 1953, **21**, 329-335.

7. IRWIN, F. W., & MINTZER, M. G. Effect of differences in instructions and motivation upon measures of the verbal level of aspiration. *Amer. J. Psychol.*, 1942, 55, 400-406.
8. LEWIN, K., DEMBO, TAMARA, FESTINGER, L., & SEARS, PAULINE. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 333-378.
9. MARKS, ROSE W. The effect of probability, desirability, and "privilege" on the stated expectations of children. *J. Pers.*, 1951, 19, 332-351.
10. ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, 31, 410-422.
11. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

Received August 17, 1955.

PERSONAL CONSTRUCTS AND PREDICTIVE BEHAVIOR¹

LEON H. LEVY

Indiana University

ACCORDING to the psychology of personal constructs as recently developed by Kelly (3), all men may be thought of as scientists in the sense that each is concerned with the prediction and control of his environment. Further, each individual seems to develop his own personal repertoire of constructs by means of which he structures his world and tries to anticipate events. These constructs may be thought of as the elements of a system by means of which the individual codifies his experience. Thus, the psychology of personal constructs is concerned with the ways in which personal construct repertoires develop and change and the ways in which they can be utilized in accounting for individual behavior.

Kelly defines a construct as a way in which two things are alike and at the same time different from a third. While more than three elements may be involved either in the development or communication of a construct, at least three must be present. The construct is considered bipolar, one pole designating the basis for the similarity and the other pole the contrast. Where persons are used as elements, some of the typical constructs found among college students are sincere vs. insincere, friendly vs. unfriendly, good looking vs. unattractive.

Among the ways in which personal constructs may be characterized, one dimension involves the extent to which the use of the construct is independent of or contingent upon the use of other constructs in the total construct repertoire. Where a construct is highly related to a large number of other constructs, it is termed *constellatory*. Where the converse is true, the construct is termed *propositional*. In Kelly's terms, "a constellatory construct is one which fixes the realm membership of its elements—for example, stereotypes: 'Anything which is a ball has got to be . . .'" "A propositional construct is one which does not disturb

the other realm memberships of its elements—for example, 'philosophical attitudes': 'Any roundish mass can be considered among other things as a ball' . . ." (3, p. 155).

If Kelly is correct about the role of personal constructs in predictive behavior, then where predictive failure has occurred, certain changes should be evident in the person's use of his constructs. The present paper is concerned with this problem in relation to the distinction between constellatory and propositional constructs.

There are two ways in which a person's use of his constructs might change. The first would consist of shifting the basis for his construction of an event or invoking a different construct or set of constructs (5). An example would be where one had made use of the construct honest vs. dishonest in construing his associates and upon finding that he was not able to predict their behavior very well, shifted to the construct reliable vs. unreliable. The second form of change in construct use is the one studied in this paper and may be referred to as reconstruction. Here the individual uses the same construct but shifts the position of the elements or reconstructs the elements. Thus, in the case of the construct honest vs. dishonest, if one construes another as honest but finds that predictions based on this construction are not supported, then one might try to raise the accuracy of his predictions by reconstructing the other person as dishonest.

Three hypotheses were consequently formulated for testing in the present study:

1. Where extensive revisions are required in an individual's predictive behavior (high invalidation), a greater amount of reconstruction will occur on constellatory constructs than on propositional constructs.

2. Where minor revisions are required in an individual's predictive behavior (low invalidation), the amount of reconstruction occurring on propositional constructs will be greater than or equal to the amount of reconstruction occurring on constellatory constructs.

3. Constellatory constructions are more

¹ This study is based upon portions of the research presented in the writer's doctoral dissertation submitted to the Ohio State University in 1954. Special appreciation is extended to Dr. George A. Kelly for his advice and guidance.

sensitive than propositional constructions to the amount of invalidation occurring for a set of predictions.

The third hypothesis is the most general of the three and is in a sense implied in the first two. All three are predicated on the assumption that constellatory constructs, by virtue of their broader range of interdependency, mediate a broader range of predictions and hence are affected only when major revisions in predictions are necessary. Because of the relative independence of propositional constructs, these will be less affected by the extensiveness of invalidation. Reconstruction is assumed to occur on propositional constructs only in connection with fairly specific or minor instances of predictive failure.

METHOD

The Role Construct Repertory Test

The Role Construct Repertory Test (RCRT) (3) was designed to elicit a set of constructs from an individual which research to date (1, 2, 4) indicates has a fair degree of stability and generality. The RCRT may be used either as an individual or as a group test. In the present study, the group form was used. In this form, *S* is first presented with a list of nineteen role titles such as, "Your mother, or the person who has played the part of your mother"; "your girl (boy) friend or closest present girl (boy) friend"; "the person with whom you usually feel most uncomfortable," etc. For each role title, *S* must supply the name of a different person. The *S* is next presented with various combinations of these names in the form of triads and is asked to indicate which two people are alike and at the same time different from the third. He is then asked to write the way in which the two are alike under the heading "Construct" and to write the term or phrase which he considers its opposite under the heading "Contrast." Each of these triads is termed a *sort*; twenty-two sorts are required of each *S*. There is no restriction on the kinds of terms used or on the repetition of terms.

Following the 22 sorts and the elicitation of constructs, *S* is finally asked to consider each of the 19 persons he has listed in terms of each of the constructs formed. For each person, he indicates whether the construct or its contrast applies. This final procedure permits the application of a nonparametric factor-analytic procedure and the subsequent identification of the relationships among the constructs used by each *S*. A complete exposition of this method may be found in Kelly's book (3).

The factorial method applied to RCRT protocols usually identifies between two and seven factors. Of these, one factor, typically the first or second extracted is saturated with a larger number of constructs than any of the others and is referred to as a general factor. In the present study the assumption was made that all constructs having significant loadings on the same

factor would be more interdependent with each other than with other constructs. Thus, constructs having significant loadings on a general factor were considered to have a wider range of interdependency as compared to other constructs. Consequently, all constructs having significant loadings on a general factor were designated constellatory; all others were propositional.

Elicitation and Classification of Constructs

RCRT's were obtained from 80 students in two sections of a course in educational psychology. Each protocol was factored until one general factor was extracted which contained eight or more constructs with significant loadings. Five constellatory and five propositional constructs were then selected from each protocol. For constellatory constructs, all those with significant loadings on the general factor were ranked in terms of the magnitude of their factor loadings, and the first five nonrepeating constructs were selected. The first five constructs appearing in the protocol which did not have significant loadings on the general factor and which contained no obvious repetitions among them were designated as propositional constructs. One additional criterion used in selecting constructs was that they had to be applicable to persons other than those on whom they were formed, i.e., they could not refer to kinship, sex, age, or occupational status. Thus, with the possible exception of the last criterion, the process of construct selection and classification was entirely a nonjudgmental one.

Out of the original set of 80 protocols, 25 were eliminated either because the general factor contained more than 17 constructs, thus not permitting the selection of the five propositional constructs, or because too few of the constructs elicited met the criteria of applicability. The participation of the 55 remaining *S*s was solicited for the experiment proper; of these, 40 (33 females and 7 males) volunteered.

Individual Sessions

Each *S* was presented with photographs of two persons. Their order of presentation was systematically alternated over *S*s. The two photographs used were of male graduate students taken when each was about 18 years of age. Students were told that the experiment was concerned with the accuracy with which personality could be judged from photographs. They were told that this was a skill which everyone possessed to some degree and that in this experiment *E* was interested in seeing how well they could utilize this skill. The *S*s were then told that *E* would be interested in how they described each person in their own terms (their constructs) and how accurately they could answer fifteen questions about each person's behavior.

For their constructions each *S* was presented with a blank on which his five constellatory and five propositional constructs, both construct and contrast, had been typed so as to represent two alternatives. The sequence of occurrence of each type of construct was the same for all *S*s with constellatory constructs located at ordinal positions 1, 3, 4, 5, and 9 in the list.

The prediction questions which *S*s were to answer were of the following type:

TABLE 1

MEAN NUMBER OF CHANGES IN CONSTRUCTION OF
CONSTELLATORY AND PROPOSITIONAL CONSTRUCTS
UNDER CONDITIONS OF HIGH AND
LOW INVALIDATION

Conditions	Construct		<i>t</i>	<i>p</i> *
	Constellatory	Propositional		
High Invalidation	1.900	1.550	2.800	.01
Low Invalidation	.775	.900	.862	ns
Difference	1.125	.650	2.387	.05

* All probabilities are for a two-tailed test since one of the problems under investigation was whether the classification of constructs as constellatory or propositional was related in any way to changes in these constructs under conditions of high and low invalidation.

1. With regard to going to parties, this person usually:
 - a. Jumps at the opportunity.
 - b. Regards them as a "bore" and goes rarely.
 - c. Will decide whether to go on the basis of who will be there.
 - d. Regards them as a "bore" but goes anyway.
 - e. Just isn't interested.
2. When he thinks about the future, he:
 - a. Is full of wonder and enthusiasm.
 - b. Has doubt about his ability.
 - c. Can see mostly calamity.
 - d. Is full of plans—has it all worked out.
 - e. Can't get very interested—more concerned with the present.

The two tasks were presented as distinct from each other and on separate sheets of paper. As *S* finished each set of constructions and predictions, *E* pretended to compare them with fictitious data on the person involved. When constructions and predictions had been made for both persons, *E* told *S* that on this trial he had only checked on the accuracy of his predictions and had not looked at his descriptions. The *E* then informed *S* that on one of the persons he had done rather poorly, below the mean in accuracy for his group (high invalidation) while on the other person he had done quite well, scoring above the mean although missing a few items (low invalidation). The *Ss* were not told either which or how many items they had presumably missed.

The *Ss* were then told that although they had not been given any more specific information than how accurate they were in their predictions, they were to try to improve on their accuracy of prediction by going through the process of prediction and description again. Their first set of constructions and predictions was then returned to them and they were given a new set of blanks. It was suggested to each *S* that although his descriptions had not been checked this time, they would be checked next along with his revised predictions, and that he might wish to change some of them. This then yielded a second set of con-

structions and predictions, in one instance following high invalidation and in the other, low invalidation. The actual order in which high and low invalidation occurred was systematically alternated from *S* to *S*.

RESULTS

From Table 1 it will be seen that the null hypothesis concerning differences in amount of reconstruction on constellatory and propositional constructs under conditions of high invalidation may be rejected at the .01 level of confidence. This provides support for the first experimental hypothesis that under conditions of high invalidation, a greater amount of reconstruction occurs of constellatory than of propositional constructs. Under conditions of low invalidation, the difference in amount of reconstruction on constellatory and propositional constructs is not significant. However, the second experimental hypothesis predicted that under these conditions, the extent of reconstruction of propositional constructs would be greater than or equal to the reconstruction of constellatory constructs. Consequently, the experimental hypothesis is not necessarily denied by these null findings. Finally, the third experimental hypothesis that constellatory constructions are more sensitive than propositional constructions to degree of invalidation is supported by the significant difference between differences in amount of reconstruction of constellatory and propositional constructs following high and low invalidation.

In each instance, then, some support was found for the predictions made regarding the relationships between amount of invalidation, kinds of constructs, and amount of reconstruction. It would thus appear that, in the sense of mediating differential predictions, both the conceptual distinction between constellatory and propositional constructs and the operations used in making this distinction are meaningful. Further, if one accepts the degree of invalidation needed to bring about a given amount of reconstruction as an indication of the susceptibility to reconstruction of that construct, the data suggest an inverse relationship between susceptibility to change and range of interdependency.

It will be recalled that the method by which constellatory and propositional constructs were distinguished was essentially a mechanical one. In order to determine whether such a procedure was necessary, or whether the distinction

TABLE 2
TYPICAL CONSTELLATORY AND
PROPOSITIONAL CONSTRUCTS

Constellatory Constructs
1. Military mind vs. few strong concepts
2. Conceit vs. unconceited
3. More reserved vs. happy-go-lucky
4. Very active vs. very lazy
5. Loquacious—easy to know vs. introvert—quiet
6. Mean to me vs. nice to me
7. Unfrivolous vs. one who "puts on"
8. Sensible vs. not too sensible
9. Happy vs. dissatisfied
10. Don't give up vs. always give up
Propositional Constructs
1. Intelligent vs. average
2. More carefree vs. more responsible
3. Understanding vs. not understanding
4. Team work vs. egotist
5. Very reserved in manners vs. happy-go-lucky
6. Self-centered vs. not quite so much
7. Very conservative vs. liberal minded
8. More serious vs. less serious
9. Always making jokes vs. dead pan
10. There's room for improvement vs. has reached goal of ethics

might be made just as well on some intuitive basis, six advanced graduate students, all of whom were familiar with Kelly's theory, were given a list of forty constructs taken at random from RCRT protocols and were asked to determine which were constellatory and which propositional. The judges were informed that half of the constructs belonged in each category. Table 2 presents a random sample of the constructs contained in this list. None of the judges was able to do better than chance. Essentially, this would suggest that whether a given term is constellatory or not depends strongly upon how the individual uses it rather than upon its semantic meaning. This functional character of constructs can only be determined by some procedure such as Kelly's which reflects individual usage.

DISCUSSION

It is possible to view the personal construct repertoire as a coding system containing a finite number of elements or constructs. If the purpose of such a system is to aid in the discrimination and anticipation of events, then one would expect that its efficiency would decline as the independence between the elements decreased.

Thus, it would be expected that the individual whose construct repertoire comprises primarily constellatory constructs would be less sensitive to the subtle nuances of interpersonal situations. For the same reason, one would expect communication to be more difficult and less efficient as the number of constellatory constructs employed by the individual increased.

These considerations raise the normative problem of the relative proportions of constellatory and propositional constructs desirable in a construct repertoire. If it is the degree of relationship between two constructs which permits the individual to formulate propositions of the form "If x then y ," e.g., "If this person is *friendly* then he *can be trusted*," then complete independence between all of the constructs in an individual's construct repertoire would leave him completely at sea in terms of making predictions and generalizations. On the other hand, if all constructs were perfectly correlated with each other, again the individual would be at a loss in dealing with a changing or probabilistic world. While neither of these extremes seems likely of occurrence, the question can be posed as to whether there is an optimal point between them.

In attempting to bring about some change in an individual's mode of response or general adjustment, the therapist or educator makes use of a variety of techniques, all of which seem to rely upon some principle of differential reinforcement. The present findings suggest that these procedures may vary in their effectiveness with the extent to which the elements involved are constellatory or propositional. It is quite probable that whenever the therapist or client ascribes failure to change as being due to "too much at stake" in the change, they are referring to a situation which has been construed in a constellatory fashion. Thus, a knowledge of the content and structure of an individual's personal construct system and a set of procedures which might reduce the constellatory nature of certain constructs would seem to be important in psychotherapy or any other situation where our aims involve changes in perception and behavior.

With respect to the latter point, many of the procedures now used by therapists may have as one of their effects the reduction of constellatory constructs. One might mention as one prominent example the technique of transfer-

ence analysis. Here the client has transferred to the therapist an entire constellation of constructions which he had formerly applied to some other important figure in his life, e.g., his father. In analyzing the client's transference reaction, the therapist may be helping the client to use such constructs as male vs. female, authority vs. nonauthority, dominating vs. not dominating, hostile vs. not hostile, arbitrary vs. not arbitrary or understanding, in a more propositional way, thus permitting him to establish role relationships with other males or authority figures where he does not continually anticipate that (and hence act as though) he will be aggressed against, punished, etc. Further research might profitably be directed toward the development and assessment of techniques aimed at reducing the constellatory nature of constructs.

SUMMARY

Starting with the assumption of the psychology of personal constructs that each individual develops and makes use of a set of constructs in attempting to predict and control his environment, the effects of apparent failure in pre-

diction were studied in relation to two kinds of construct (constellatory and propositional) differentiated on the basis of their range of interdependency. Three specific hypotheses were tested and each found experimental support. From these findings, it might be concluded that there is an inverse relationship between the range of interdependency of a construct and its susceptibility to change following predictive failure. Several possible implications of these findings are discussed.

REFERENCES

1. HAMILTON, R. JANE. Generality of personal constructs. Unpublished doctor's dissertation, Ohio State Univer., 1952.
2. HUNT, D. E. Studies in role concept repertory: Conceptual consistency. Unpublished master's thesis, Ohio State Univer., 1951.
3. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
4. SHOEMAKER, D. J. The relation between personal constructs and observed behavior. Unpublished master's thesis, Ohio State Univer., 1952.
5. POCH, SUSANNE M. A study of changes in personal constructs as related to interpersonal prediction and its outcomes. Unpublished doctor's dissertation, Ohio State Univer., 1952.

Received August 17, 1955.

HUMOR AND ANXIETY

JOHN DORIS AND ELLA FIERMAN

Yale University

FOLLOWING Kris' lead (5, 6) various psychoanalytically oriented writers (4, 7, 8, 10) have pointed out that Freud's position as stated in *Wit and Its Relation to the Unconscious* implies a relationship between humor and anxiety. These writers have emphasized the aggressive and sexual, as well as the nonsensical, aspects of humor. This literature raises the experimental question as to whether subjects differentiated on the basis of a self-rated anxiety scale respond differently to humorous stimuli.

METHOD

A self-rated general anxiety questionnaire¹ was administered during regular class sessions of introductory psychology courses to 419 Yale men and 197 Connecticut College women.² From the distributions of scores on this questionnaire, for each of the respective colleges, extreme groups of high anxious (HA) and low anxious (LA) subjects (Ss) were selected. The Yale Ss were selected from within the upper and lower 6 per cent of their distribution and the Connecticut Ss were selected within the upper and lower 9 per cent. There were 56 Ss in all, 28 HA and 28 LA. Each anxiety group contained equal numbers of men and women. The Ss were not told that their selection for participation in the humor section of the study was based upon the anxiety questionnaire that they had filled out several months previously. Participation in the humor study was on a volunteer basis but of all students approached only three refused to serve.

The humor stimuli were cartoons selected from the Mirth Response Test (10) which consists of cartoons from *The New Yorker* and similar magazines.³ The cartoons of the Mirth Response Test were given to a group of 15 psychiatrists, psychologists, and psychiatric social workers who rated each cartoon on the amount of aggression, sex, and nonsense in the cartoon. The ratings were made on graphic scales which were descriptively labeled to indicate increasing amounts

of the characteristic being rated, and subsequently scored in equal linear units from 1 through 9. The ratings were then averaged for each cartoon and three groups of six cartoons each were selected from the original 31, each composed of cartoons in which either the sexual, aggressive, or nonsensical aspect of humor was respectively predominant. Thus, the six aggressive cartoons received a mean rating of 7.8 on aggression, 1.4 on sex, and 5.9 on nonsense. The sexual cartoons were rated 3.6 on aggression, 7.0 on sex, and 5.4 on nonsense. Ratings on the nonsensical cartoons were 3.2 on aggression, 1.4 on sex, and 6.8 on nonsense.

In the experimental situation the cartoons were presented to *S* one at a time in a standardized order that intermingled the various content categories of aggression, sex, and nonsense. The *S*'s vocal and facial responses to the cartoons were rated by the examiner on a 6-point scale from disapproval to pronounced pleasure.⁴ After the *S* had examined all cartoons he was asked to go over them again indicating his degree of preference on a graphic rating scale extending from Very Much Disliked to Very Much Liked. The *S* was then asked to describe the point of each joke. The examiner (*E*) wrote down this description and it was later scored for comprehension of the cartoon.

There were two examiners, one male and one female. Each examiner tested 28 Ss: 7 men and 7 women from each of the anxiety groups. An effort was made to select the Ss in such a fashion that at the time of testing the examiner would not know whether the *S* being tested belonged in the HA or LA group. With the exception of four cases this was possible.

RESULTS

The Ss' explanations of the cartoons were first scored for comprehension according to criteria established on the basis of the investigators' previous experience with the cartoons. These criteria, with minor modifications, were similar to those recommended by the originators of the Mirth Response Test. The investigators were able to score independently a sample of 144 cartoon interpretations with 95 per cent agreement.

Scoring of the cartoon interpretations revealed that approximately 85 per cent of the Ss

¹ This questionnaire is a revision of the Questionnaire on Adult Forms of Anxiety and Worry used by Gordon and Sarason (2). Copies may be obtained from the present authors upon request.

² The authors wish to express their thanks to Dr. S. B. Sarason of Yale and Dr. M. Applezweig of the Connecticut College for Women for their cooperation in conducting the experiment at their respective schools.

³ The authors would like to express their thanks to Dr. F. C. Redlich and Dr. J. Levine of the Yale Psychiatry Department for their permission to use materials from the Mirth Response Test and for their generous cooperation in making this study possible.

⁴ In an attempt to estimate the reliability of the mirth-response ratings, six Ss other than those used in the experiment were asked to look through the 31 cartoons of the Mirth Response Test while the authors made independent judgments of their mirth responses. The reliability coefficients of the authors' judgments for these six Ss were respectively .80, .76, .90, .45, .74, .83.

TABLE 1
MEAN PREFERENCE SCORES FOR CARTOONS BY
ANXIETY GROUPS
($N = 28$ in each group)

	Nonsensical Cartoons	Aggressive Cartoons	Sexual Cartoons	All Cartoons
High Anxious	8.7	9.2	8.0	8.7
Low Anxious	9.3	10.2	8.2	9.4
t^*	1.59	2.25	0.35	1.95
p	.20 > < .10	.05 > < .02	—	.10 > < .05

* Tests for homogeneity of variance were done for all t tests reported in this study. In no case was rejection of the hypothesis of a common population variance indicated.

TABLE 2
MEAN PREFERENCE SCORES FOR CARTOONS BY ANXIETY GROUPS SUBDIVIDED
ACCORDING TO EXAMINER-SUBJECT SEX COMBINATION
($N = 14$ in each group)

	Nonsensical Cartoons		Aggressive Cartoons		Sexual Cartoons		All Cartoons	
	Opposite Sex	Same Sex	Opposite Sex	Same Sex	Opposite Sex	Same Sex	Opposite Sex	Same Sex
High Anxious	8.4	9.0	9.2	9.3	7.7	8.2	8.6	8.9
Low Anxious	9.5	9.1	11.0	9.4	8.2	8.2	9.7	9.0
t	2.09	—	2.86	—	0.56	—	2.44	—
p	.05 > < .02	—	.01 > < .001	—	—	—	.05 > < .02	—

misunderstood two of the sexual cartoons.⁵ It was decided to remove these cartoons before scoring the preferences of the Ss.⁶ The point of the joke being missed, it seemed invalid to compare Ss on their appreciation of the humor; and since Ss in HA and LA and male and female categories had equal difficulty in comprehension, it was felt that there was no connection between failure to understand these cartoons and the major variables and categories of interest. For the remaining cartoons there were only a few scattered cases of misinterpretation equally divided between groups, and for purposes of further analysis these few instances were treated exactly as the rest of the data. It was assumed, and inspection of the data lent

⁵ A Cobean cartoon with his familiar "undressing" theme and a Kovarsky cartoon satirizing this theme. These cartoons required certain knowledge that the S group may not have had at its disposal. Understanding of the Cobean cartoon required that one be able to recognize the appearance of an oriental eunuch. The Kovarsky cartoon required that one have some familiarity with the Cobean "undressing" theme.

⁶ The mean ratings by psychiatrists, psychologists, and social workers for the four retained sexual cartoons were 7.84 on aggression, 7.08 on sex, and 6.84 on nonsense.

support to the assumption, that the preference scores of the Ss in these few instances did not materially distort the means of the groups.

For each of the cartoons retained for further analysis, S had a rating from 0 to 15 representing his preference on the graphic scale and a rating from 0 to 5 based on E 's judgment of his mirth response. The respective means of these ratings on each set of sexual, aggressive, and nonsensical cartoons, and on all 16 cartoons combined, were considered to be S 's preference and mirth-response scores. Comparisons be-

tween the anxiety groups were then made on the bases of these scores.

In terms of the average preference scores (Table 1) the HA group showed a tendency ($.10 > p > .05$) to give lower average ratings than the LA group on all types of cartoons combined. On the aggressive cartoons the preference scores of the HA group were lower ($.05 > p > .02$) than those of the LA group. There were no reliable differences between the groups on the nonsensical and sexual cartoons.

Closer examination of the data revealed, however, that not all members of the HA and LA groups were contributing equally to the obtained differences. Rather, it was those Ss tested by an examiner of the opposite sex that contributed the major part of the difference. The Ss were therefore classified not only according to their anxiety grouping but also according to the conditions of administration. This analysis is presented in Table 2. In the *opposite* condition the male E administered cartoons to female Ss and the female E administered cartoons to male Ss. In the *same* condition the male E administered cartoons to male Ss and the female E administered cartoons to the female Ss. It is

TABLE 3

MEAN MIRTH-RESPONSE SCORES FOR CARTOONS BY ANXIETY GROUPS WITH EXAMINER AND SUBJECT OF OPPOSITE SEX

($N = 14$ in each group)

	Nonsensical Cartoons	Aggres- sive Car- toons	Sexual Car- toons	All Car- toons
High Anxious	3.30	3.40	3.51	3.39
Low Anxious	3.83	3.79	3.53	3.74
t	2.18	1.28	—	1.46
p	.05 > .02	—	—	—

now noted that differences between HA and LA Ss exist only in the *opposite* condition for the nonsensical ($.05 > p > .02$), aggressive ($.01 > p > .001$), and all cartoons combined ($.05 > p > .02$). No significant difference is found in the *same* condition for nonsensical, aggressive, and all cartoons, nor in either condition for sexual cartoons.

Mirth-response scores were treated in the same manner as preference scores. There were no differences between total HA and LA groups in the average mirth-response scores. Since analysis of the data according to conditions of administration proved enlightening for the preference scores, however, a similar analysis was carried out for the mirth-response scores, which revealed a rather similar pattern to that obtained with the preference scores. There were no differences between the anxiety groups in the *same* condition of administration while there was some tendency of the HA group to score lower than the LA group in the *opposite* condition. However, only in the case of the nonsensical cartoons ($.05 > p > .02$) was the difference significant. Table 3 reproduces only the results for the *opposite* condition for the mirth-response scores. All differences for the *same* condition were negligible.

DISCUSSION

Two results seem to stand out in this experiment. First, for this group of Ss, there seems to be a relationship between S's rating of his susceptibility to anxiety as determined by a general anxiety scale and his preference rating for cartoons of aggressive content. Second, this relationship depends upon the social context of the humor stimuli; i.e., results depend upon which S is tested by which E.

This interrelationship of anxiety, preference rating, and social context also seems to be present in the case of nonsensical cartoons, although evidence is not as clear-cut as with the aggressive cartoons. In the case of the sexual cartoons, there was no differentiation of HA and LA groups under any condition. It may be that the smaller number of sexual cartoons made for less reliable measurement. Further study is necessary to explain this failure of the sexual cartoons to differentiate.

These results suggest that further study of the relationship between humor and anxiety may be profitable. They strongly indicate, however, that such experimentation should not focus exclusively on the stimulus materials, i.e., jokes or cartoons, per se. The social context appears to be crucial. In this, the results parallel those of Perl (9), who demonstrated that Ss rated jokes differently when tested alone than when tested in a group. In our study, the personalities of the Es must have provided cue differences which altered the situation for S. Standardized administration procedures and E's lack of knowledge concerning the anxiety classification of Ss during testing suggest that these cue differences were not the result of E's consciously altering his behavior for the different groups of Ss.⁷ While the exact nature of these cue differences cannot be determined on the basis of this experiment, the sex difference would seem most immediately obvious to the Ss, and appears worth further investigation. Such an investigation would, of course, require a more representative sample of experimenters than the design of this study permitted (3).

That the mirth-response score did not discriminate as well as the preference score is of interest in view of the fact that other investigators (7) found a mirth-response score adequate for discrimination within a patient population of HA and LA Ss who were selected on the basis of E's estimate of their anxiety level while in the test situation. Further study would seem to be indicated to determine the relative merits of these two scores in humor studies, as well as the alternative techniques of selecting anxiety groups.

⁷ During the test administration the examiners attempted to classify the Ss they saw as either HA or LA. Their success in this classification did not exceed chance expectations.

REFERENCES

1. FREUD, S. Wit and its relation to the unconscious. *The basic writings of Sigmund Freud*. New York: Modern Library, 1938.
2. GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, **23** (3), 317-323.
3. HAMMOND, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51** (2), 150-159.
4. JACOBSEN, EDITH. The child's laughter, theoretical and clinical notes on the function of the comic. *Psychoanal. Stud. Child*, 1946, **2**, 39-60.
5. KRIS, E. Ego development and the comic. *Int. J. Psychoanal.*, 1938, **19**, 77-90.
6. KRIS, E. Laughter as an expressive process. *Int. J. Psychoanal.*, 1940, **21**, 314-341.
7. LAFFAL, J., LEVINE, J., REDLICH, F. C., & FIERMAN, ELLA. The relationship between anxiety and responses to humorous cartoons dealing with hostile aggression. Paper read at East. Psychol. Assoc., Boston, April, 1953.
8. LAFFAL, J., LEVINE, J., & REDLICH, F. C. An anxiety-reduction theory of humor. *Amer. Psychologist*, 1953, **8**, 383. (Abstract)
9. PERL, RUTH. The influence of a social factor upon the appreciation of humor. *Amer. J. Psychol.*, 1933, **45**, 308-314.
10. REDLICH, F. C., LEVINE, J., & SOHLER, T. P. A mirth response test: preliminary report on a psychodiagnostic technique utilizing dynamics of humor. *Amer. J. Orthopsychiat.*, 1951, **21**, 717-734.
11. REICH, ANNIE. The structure of the grotesque-comic sublimation. *Bull. Menninger Clin.*, 1949, **13**, 160-171.

Received August 22, 1955.

FACTORS IN THE SERIAL RECALL OF NAMES OF ACQUAINTANCES¹

RUE L. CROMWELL

George Peabody College for Teachers

THE studies reported here examine from a novel perspective the influence on serial recall of several factors, most of them familiar. Conventionally, the process of recall is investigated either by experimentally varying the conditions of learning and reinstatement, on the one hand, or by soliciting the uncontrolled recollection of life experiences, on the other. The present studies attempt to preserve the advantage of the more naturalistic kind of investigation in which the learning underlying recall has occurred in the normal course of life experience, while retaining a degree of control more characteristic of experimental studies. In each case, the task set the subject involves recalling the names of acquaintances that fit certain specifications. The order of recall is noted. After recall, the subject is requested to *rank* the names with respect to the variables under investigation.

In the first study, five variables are examined: frequency, recency, and pleasure, from associationism and hedonistic theories; degree of acquaintance (on common-sense grounds); and a factor drawn from the personal construct theory of Kelly (3). The second study serves a dual function. While it is primarily designed to check findings obtained in the previous study, it also permits a more extended analysis of the pleasure variable.

STUDY I

The hypotheses underlying the first study stated in terms of the recall-ranking situation, are as follows:

1. *Pleasure hypothesis*: The order of recall is positively correlated with the order in which the recalled elements are ranked in terms of pleasure. The rationale for this hypothesis can be drawn from the hedonistic formulations of psychoanalysis. According to Freud there is "the working of a tendency to ward off from memory that which is unpleasant" (2, p. 68).

¹ This article is based on a master's thesis at The Ohio State University and a follow-up experiment. Advice and criticism on various phases of the study were received from George A. Kelly, Paul Mussen, Richard Lundy, Robert J. Wherry, Raymond Norris, and Rupert Klaus.

More "pleasant" objects, by this hypothesis, should be recalled early in a recall sequence, less "pleasant" ones later on, and very "painful" ones "repressed" and therefore unavailable to immediate recall.

2. *Frequency hypothesis*. The order of recall is positively correlated with the frequency of contact with the recalled elements.

3. *Recency hypothesis*. The order of recall is positively correlated with the recency of contact with the recalled elements.

4. *Acquaintance hypothesis*. The order of recall is positively correlated with the degree of the S's acquaintance with the recalled elements. This hypothesis is offered on the common-sense grounds that people recall best what they "know" best.

5. *Construction hypothesis*. The items recalled adjacently (in terms of sequence) are grouped by the S as similar to each other to a degree better than chance. The personal construct theory of Kelly (3) suggests that when an individual recalls a number of items from past experience, he recalls them in groups as a function of his own manner of classification or construction. When the recalled items are names of people, an S might recall college acquaintances, relatives, neighbors, and then teachers. In this way, items adjacent to each other in sequence of recall may tend to be grouped by the S as similar in contrast with nonadjacent recalled items.

Method

Subjects. Five elementary psychology classes, 151 students in all, were used in Study I. The classes, approximately equal in size, were respectively assigned to the five experimental groups. Through failure to recall the eight names required in the instructions, 14 cases were discarded.

Materials. Each S was given a dittoed data sheet form and eight blank 2½" x 3" cards. A blackboard was used to illustrate the instructions to each group.

Instructions. The instructions, read verbatim to each of the five experimental groups, gave directions to perform the following tasks.

1. The Ss were told they would be asked to recall, and write down in order as they recalled, the names of eight living acquaintances. Each of these acquaintances was to have the same first name, which would be

TABLE 1

MEAN RANKINGS FOR THE FOUR RANKED VARIABLES

Ranking	P	F	C	A
First	3.36	3.39	3.58	3.20
Second	3.88	3.71	4.28	4.05
Third	4.31	4.45	4.10	4.39
Fourth	4.38	4.62	4.56	4.33
Fifth	5.11	4.55	4.34	4.89
Sixth	4.96	4.77	4.80	4.62
Seventh	4.93	5.14	5.11	5.41
Eighth	5.07	5.36	5.23	5.11

Note.—P stands for *pleasure*; F, *frequency*; C, *recency*; and A, *acquaintance*.

TABLE 2

INTERCORRELATIONS AMONG THE RANKED VARIABLES AND THE ORDER OF RECALL

	R	P	F	C	A
Order of recall (R)	x				
Pleasure (P)	.91	x			
Frequency (F)	.96	.89	x		
Recency (C)	.94	.82	.91	x	
Acquaintance (A)	.90	.93	.91	.88	x

specified by the experimenter. The *S*'s were to write each acquaintance's name on a blank card, which was numbered on the reverse side to indicate the order of recall. As soon as *S*s indicated they had understood these instructions, a common American first name was announced and they were allowed 20 minutes to recall and write down the names of eight people with this first name. Group I was given "John;" Group II was given "Mary;" Group III, "James" or "Jim;" Group IV, "Jane;" Group V, "William."

The order of the following steps (2 to 6) was systematically varied from group to group. The order given below was presented to Group I only.

2. The *S*s were asked to rank the eight recalled names in terms of "how well you like the persons you have recalled." The *S* then recorded this ranking on his data sheet by referring to the card number in terms of order of recall.

3. The *S*s were then asked to rank the recalled persons "in terms of how much contact you have had with these persons during your life." This ranking was likewise recorded on the data sheet.

4. Next, the *S*s were asked to rank and record the recalled names "in terms of how long it has been since you have had contact with each person," the most recent being ranked first.

5. Then, the *S*s were asked to rank and record the cards "in terms of how well you know each person," the best known person being first.

6. The *S*s were then told: "Group the cards into three groups in terms of how they are alike in some way. Put any number in each group. What basis you have for the distinction is up to you." The group memberships were recorded with respect to the initial order of

TABLE 3

PARTIAL CORRELATIONS BETWEEN THE RANKED VARIABLES AND THE ORDER OF RECALL

Variable	Partial correlation with the order of recall	Significance level (<i>df</i> = 3)
Pleasure	$r_{pr.fca} = .61$	$p > .20 < .30$
Frequency	$r_{fr.pea} = .61$	$p > .20 < .30$
Recency	$r_{er.pfa} = .68$	$p > .20 < .30$
Acquaintance	$r_{ar.pfe} = -.45$	$p > .40 < .50$

recall for each name. The "bases for the distinctions" were also written on the data sheet. At this point the *S*'s data sheet, now complete, contained four rankings and one grouping.

Controls. The "first name" stimulus word and its designated sex were varied from group to group. In order to reduce or compensate for ranking habits, the order of the ranking tasks was systematically varied from group to group, the recalled names were put on separate cards, the cards were numbered on the reverse side only, and they were shuffled by the *S*s after each ranking and grouping was performed.

Proper names were used as items for recall because they conveniently fit the procedure. The *S* was not allowed to use his own name or the names of deceased persons, to avoid cultural influences demanding acceptable ways of reacting to self and deceased.

Results

The rankings on the several variables were first summated throughout the 137 *S*s, and then mean ranks were computed. A single table of mean ranks (Table 1) was therefore obtained, rather than the many individual ranks.

To evaluate the magnitude of the trends apparent in Table 1, product-moment correlations were computed between the mean ranks for each ranked variable and the (mean) order of recall.² Each of the four correlations, presented in the first column of Table 2, was significantly above zero with probability levels beyond the .01 level. The reader is reminded that these correlations describe the smoothed-out group trend and are consequently much higher than individual rank-difference correlations would have been.

Since the ranked variables correlate with each other as well as with the order of recall (Table 2), partial correlations were computed

² It was found that reranking the mean ranks and also scaling the ranks for normality brought correlations which were not appreciably different from the product-moment correlations.

TABLE 4
EXPECTED AND OBSERVED NUMBER OF PEOPLE
NAMING "ISOLATES"

Number of isolates	Expected frequency	Observed frequency
0	2.01	4
1	9.10	10
2	19.94	24
3	22.19	20
4	35.32	34
5	12.11	9
6	28.26	23
8	8.07	13
	$N = 137.00$	$N = 137$

(Table 3). *Pleasure*, *recency*, and *frequency* retain a positive relationship with the order of recall. *Acquaintance* becomes negative. However, none of these partial correlations was significant at the .05 level of confidence.

With the *construction* hypothesis, the problem was to determine the degree to which adjacent objects in the sequence of recall were grouped together as similar. An object that was grouped as similar to neither of the adjacently recalled objects would be directly contrary to the hypothesis. Such negative instances, which may be called "isolates," facilitated the testing of the hypothesis in reciprocal form: that the number of isolates per *S* is fewer than would be expected by chance.

Since each of the three groups could have any number of items in its membership, it was necessary to derive the expected frequencies for all the possible arrangements of order and classification of the items. These various arrangements were in turn classified according to the number of isolates in each. Finally, the expected number of isolates per individual could be compared with the observed number of isolates per individual. (For a full account of this derivation, see 1, p. 42-46.) Table 4 shows the expected and observed number of people having isolates at each possible frequency level. The most convenient and favorable dichotomization by which to test the hypothesis with one-by-two chi square divides the group between the two- and three-isolate categories in Table 4. With this division point, the chi-square test shows a trend which falls short of significance. ($\chi^2 = 2.04$; $.20 > p > .10$).

STUDY II

The mean data for the four ranked variables in the experiment just described are very similar to each other (see Table 1). The possibility is raised, therefore, that some extraneous variable is exerting an influence common to all these data. For example, extraneous ranking habits, rather than the variables actually being studied, could have led to the similarly positive findings.

On the assumption of such a spurious relationship it would be expected that an assumed neutral variable, such as *height*, would also be positive because of the extraneous influence. In this study the *pleasure* variable was compared with ranking in terms of *height*. It was predicted that with the same group of subjects the *pleasure* variable would have a significant relationship to the order of recall and the *height* variable would not.

The second aspect of this study is based on an observation made in the first study. During the recall period in the first study the subjects were observed to write down two or three or four names immediately and quickly. After this, they paused for great lengths of time until all the eight names were recalled. These observations of the recall behavior were ignored by the experimenter until the data of the first experiment were analysed. It was then noted that differences occurred between the rankings of the later and earlier recalls. As can be seen in Table 1, the *pleasure* trend levels off, the *recency* and *frequency* trend becomes more regular, and the *acquaintance* trend becomes more irregular. Thus, the possibility arose that these changes in the trends of the ranking might be related to the change from the immediate, tip-of-the-tongue recall to the slower, harder-to-recall names.

The study in its second aspect is therefore concerned with one of these rankings, the *pleasure* ranking, and its relation to the availability of the recalled items. Since the curve for *pleasure* levels off in the functional relationship with the order of recall, it is hypothesized that the order of recall before the initial pause has a greater relationship to the *pleasure* ranking than has the order of subsequent recall.

Method

Subjects. A group of 35 girls from freshman psychology classes took part in the experiment as a part of

course requirements. Ten of the *Ss* were rejected because of failure to recall the eight names required. One *S* failed to indicate the point of first pause.

Instructions. The instructions given the *Ss* before the recall period were identical to those in the first experiment, except that only the *pleasure* and *height* variables were ranked. The *Ss* were asked to recall and write down the names of eight living acquaintances whose first name was John. The names were put on individual cards with the order of recall indicated by numbering on the reverse side. The period for recalling names was approximately 15 minutes. At the end of the recall period, *Ss* were asked to indicate on their data sheet the number of names which they recalled immediately without having to pause. The *Ss* were then asked to shuffle the cards and then rank them in terms of how well they liked the individuals they had recalled. The ranking was recorded on the data sheet. The *Ss* were finally asked to rank the cards according to the height of the persons named. Again, the ranking was recorded on the data sheet.

Results

The mean ranking of *pleasure* correlated $+.60$ with the order of recall, significantly greater than zero. Between the mean ranking for *height* and the order of recall the correlation was $+.18$, not a significant degree of relationship. The correlation between *pleasure*-ranking and order of recall was not as great as that reported in the first experiment, presumably because the *N* here was smaller. It was greater than the correlation between *height* and order of recall to a probability level of $.06$. Thus, the first hypothesis was confirmed.

The point of first pause, as reported by the *S*, was used to split the *pleasure*-ranking of the eight names into two groups. The first group contained the names recalled before the first pause. The second group contained the names recalled after the first pause. Because the point of pause varied from *S* to *S*, the sizes of the two groups also varied from *S* to *S*. Table 5 shows the distribution of the *Ss* in regard to the point of pause.

By separating the group of eight names at the various points indicated in Table 5, the group sizes for an individual *S* was either one and seven, two and six, three and five, or four and four. The original ranking of the eight names was reranked within these two smaller groups. Rank-difference correlations were then computed from each of the smaller groups. (An exception existed in the one case with the one-seven split of the eight names. Here, with only one set of scores, a correlation among the early recalls was impossible.) This procedure yielded

TABLE 5
DISTRIBUTION OF THE POINTS OF FIRST PAUSE IN
RECALLING EIGHT NAMES

Occurrence of first pause	Frequency
Before 1st name	0
Between 1st and 2nd name	1
Between 2nd and 3rd name	5
Between 3rd and 4th name	10
Between 4th and 5th name	4
Between 5th and 6th name	2
Between 6th and 7th name	2
Between 7th and 8th name	0
	$N = 24$

two rank correlations for each *S*, one describing the relationship between *pleasure* and order of recall before the initial pause, the other describing the relationship between *pleasure* and the order of recall after the initial pause.

A sign test was used to test the second hypothesis. For each *S* the chance expectancy that one rank correlation exceeds the other is $.50$. Among the 24 cases, 18 were in the direction of the hypothesis, with the correlation based on the earlier recalls greater than that based on the later recalls. Four cases were in the direction opposite the hypothesis. One *S* had equal correlations and, therefore, was counted as contrary to the hypothesis. The remaining case, noted in the preceding paragraph, was not subject to rank-correlation comparison because of the one-seven split. If the latter case is ignored, the final count would be 18 positive observations out of 23. If the latter case is included on the negative side, the final count would be 18 out of 24. In either instance, the sign test indicates significant results in the predicted direction at the $.01$ level of probability.

Discussion

The first four hypotheses in the first study were substantiated by the zero-order correlation to a high degree. How well the *Ss* liked a person, how much contact they had had with him, how recently they had seen him, and how well they "knew" him were all directly related to how promptly they recalled his name.

However, the partial correlations between order of recall and *pleasure*, *recency*, *frequency*, and *acquaintance*, respectively, fell short of significance. It is to be noted here that the test of

significance assumes in each case that eight pairs of observations were made rather than 137. In general, the partial correlation findings suggest that much of the same thing is being measured in each of the correlations.

The *construction* hypothesis had the underlying assumption that people recall names categorically, or in groups, as a function of the way they conceptualize their past interactions. The hypothesis was tested by having the *S* regroup the names after the recall period and use specifically three groups. This method brought a nonsignificant trend in the predicted direction ($.20 > p > .10$).

In the second study the earlier recalls in a sequence were shown to have a greater relationship to the pleasure ranking than the later recalls. This finding rules out the possibility of a linear relationship between the two variables. The curve of pleasure ranking as a function of order of recall levels off, as is suggested in Table 1.

Whether the reported first pause, here used, is the most appropriate point at which to divide the earlier from the later recalls remains to be established. Presumably, the most useful division would be the one yielding the sharpest difference in relationships when comparing the earlier recalls with the later ones. In the present study it seemed reasonable to the experimenter, and communicable to the *Ss*, to separate those few immediately available recalls, which required no pause, from the later, more difficult recalls.

In more general terms, the results of these studies seem to have contributed to the knowledge of serial recall in a controlled association situation. The order in which the things are recalled, as a person reflects upon his past interactions in response to instructions, is to some extent predictable. The results also suggest that, for the time being, no single theory of serial recall need be accepted to the exclusion of other theories. Different theories and different points of view were employed equally well in predicting the present results. Moreover, a new kind of methodological approach has been suggested. A refinement of the five variables studied here and an exploration of other variables might afford a still better understanding of this temporal aspect of verbal behavior.

SUMMARY AND CONCLUSIONS

In the first experiment, 137 subjects in five groups were given a first name and were asked to recall eight living acquaintances who had this name. They were then asked to rank the recalled names as to (a) how well they like each person, (b) how frequently they have had contact with him, (c) how recently they have had contact with him, and (d) how well they know him. A fifth task was to put the recalled names into three groups so that each group was alike and at the same time different from the other groups. These five tasks yielded rankings referred to as the *pleasure*, *frequency*, *recency*, *acquaintance*, and *construction* variables, respectively. The order of tasks was varied among the five groups.

1. Rankings on the first four tasks were highly correlated with the order in which the names were recalled.

2. On the grouping task, order of recall fell short of a significant relationship with the constructive groupings.

In the second experiment, 24 subjects in one group were given a first name and were asked to recall eight living acquaintances who had this name. When the recalls were made, the subjects designated the point in the series at which they had to pause for the first time in order to recall more names. After this, they were asked to rank the recalled names in terms of how well they liked each person, and of the estimated heights of the individuals.

1. The *pleasure*-order of recall correlation was significantly greater than zero, and the *height*-order of recall correlation was not.

2. The *pleasure*-order of recall correlation was greater than the *height*-order of recall correlation at a probability level of .06.

3. The order of the recalls made before the initial pause had significantly greater relationship to the *pleasure* ranking than had the order of the later recalls.

REFERENCES

1. CROMWELL, R. L. Constructive factors in serial recall. Unpublished master's thesis. The Ohio State Univ., 1952.
2. FREUD, S. *A general introduction to psychoanalysis*. New York: Garden City, 1943.
3. KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.

Received August 22, 1955.

REPLICATION AND COMMENTS: "AN EXPERIMENTAL REUNION OF PSYCHOANALYTIC THEORY WITH PERCEPTUAL VIGILANCE AND DEFENSE"

CHARLES D. SMOCK

Iowa Child Welfare Research Station, State University of Iowa

BLUM (2) has recently obtained experimental results that are interpreted as lending support to the hypothesis that perceptual vigilance and defense phenomena are both consequences of repressed impulses. The experiment is of general theoretical interest since the obtained results represent perceptual selectivity phenomena which cannot be interpreted purely in terms of frequency or familiarity of the stimulus (4, 5, 10, 11). On the contrary, the findings would seem to support the contention that motivational states have direct perceptual consequences.

The experimental hypothesis was derived from the psychoanalytic premise that repressed aggressive and sexual impulses are continually striving for outlet in conscious behavior. On an unconscious or preconscious level individuals should therefore be responsive to environmental cues relevant to the gratification of these potentially dangerous impulses (vigilance). As the impulses impinge on the ego, however, anxiety is aroused evoking the ego-defensive mechanisms (learned reactions to anxiety), thus avoiding the anxiety producing cues (defense). "In sum we are led by psychoanalytic theory to postulate two opposing tendencies—vigilance and defense—and to predict the exact operation of each by controlling the level of awareness" (2, p. 94).

Seven males and seven females served as Ss in the Blum experiment. The Ss were presented sets of four miniature Blacky pictures on a card in a mirror tachistoscope. Two of these pictures, Oral Sadism (OS) and Masturbation Guilt (MG), served as experimental stimuli. The remaining two, Oral Eroticism (OE) and Identification Process (I), were "distractors" or control stimuli. Level of awareness was defined in terms of the exposure time of the stimuli; .03 seconds corresponding to a "low level of awareness" and .20 seconds to a higher level of awareness. A base series, to obtain the level of responsiveness presumably determined by the physical properties of the stimuli, was first administered

to the Ss. This consisted of 54 trials at .03 seconds with instructions to indicate which one of the four pictures "stands out" the most clearly. The Ss were then shown two regular Blacky cards (MG and OS, stimuli relevant to the repressed impulse) for the sensitization experience. The sensitization instructions were intended to structure one picture as the "traumatic" stimulus for each S (arousal of the repressed impulse) and one as the neutral stimulus.

The Ss were then administered 54 additional exposures to the cards at .03 seconds exposure time, again indicating on each trial which of the four stimuli appeared to "stand out" (vigilance series). The vigilance prediction was tested by comparing the ratio of traumatic/neutral choices in the "base" series with the same ratio in the vigilance series. Eleven Ss (six males and five females) shifted in the direction of choosing relatively more frequently the "traumatic" stimuli. The number of shifts is significant at the $<.05$ level¹ of confidence as evaluated by the sign test.

Following the presentation of the vigilance series the Ss were exposed to the stimuli under conditions necessary for testing the defense hypothesis. The Ss were informed that the two experimental pictures would now be present on each of the cards; their task was to locate the correct position of the designated picture. The Ss were instructed to locate the "traumatic" picture for 6 trials; then the neutral for 6 trials until each S had been given 36 such trials for each picture. The ego was "brought into play" further by exposing the stimuli for longer (.20 seconds) duration periods (higher level of awareness). The defense hypothesis was tested by comparing the number of correct locations of the traumatic and neutral stimuli. The results conformed to the prediction that Ss would make more errors in locating the traumatic picture (12 of 14; $p = .006$).

¹The level of confidence reported was obtained using a one-tailed test.

The fact that Ss after the sensitization experience were more responsive to the critical stimulus presented at low exposure periods and less accurate in locating the critical stimulus at increased periods of exposure was interpreted as lending support to the psychoanalytic assumptions regarding repressed impulses (2, p. 97). This paper presents experimental evidence that failure to control certain stimulus variables in the test of the defense hypothesis preclude unequivocal interpretation of these data as a function of a perceptual defense mechanism. Brief consideration will also be given to the possibility of alternative interpretations of the data obtained under these experimental conditions.

It is correctly stated by Blum (2, p. 95) that in order to attribute the changes in perceptual selectivity to the operation of vigilance and defense mechanisms it is necessary to determine the level of responsiveness to the test stimuli purely as a function of the physical properties of those stimuli. A further requirement was that the four stimuli should elicit approximately equal frequencies of response in the control series (3). The pre-experimental base series presumably served as the appropriate control for these stimulus factors in both the vigilance and defense series. However, it should be noted that the two situations present different problems of experimental control. The test of the vigilance hypothesis involves a comparison of ratio scores which takes into account the individual's presensitization level of response to the critical and neutral stimuli; i.e., a statistical control for stimulus factors is employed. The relative frequency of response to the four stimuli in the base series (3) would also be an important consideration for providing a *range* of potential vigilance effect and for the application of certain statistical tests to the obtained scores.

The *distribution* of these pre-experimental responses, however, is most crucial to the interpretation of the defense effects. The test of this hypothesis is based on a difference score derived from a comparison of the frequency of correct location of the critical and neutral stimuli in independent series of trials. The scores could be interpreted as reflecting defense effects only if it has been demonstrated that the two stimuli have equal probabilities of eliciting the correct response (i.e., equal ease of locating) prior to the experimental sensitiza-

tion experience. A large difference score (defense) could otherwise be purely a function of the distinctiveness of a particular stimulus as compared to one or more of the other three stimuli on the card. In order to interpret the results in terms of the defense hypothesis it is therefore necessary to assume that relatively equal frequency of response to the four stimuli in the base series (.03 seconds exposure) is adequate control for the differential ease of locating the stimuli in the defense series (.20 seconds). Consideration of the results from an initial attempt to replicate the experiment suggests this assumption is not tenable.

EXPERIMENT I

The procedures were as similar as possible to the original experiment with the exception that the control stimuli (i.e., "distractors," I and OE) were not altered. The Identification picture (I)² consists of two distinct forms (a large and a small dog). If unaltered, the larger form is quite similar structurally to the critical picture for the female Ss (OS). The mean number of responses to the I picture was 23.79 (base series) and 22.29 (vigilance); whereas the means of the other pictures for the base series were 13.96 (OE), 11.92 (OS), and 14.29 (MG), and for the vigilance series, 11.21, 13.17, and 16.67 respectively. The distinctiveness of the I picture at .03 seconds exposure resulted in clearly asymmetrical responsiveness in the base series. The obtained results yield statistically significant vigilance effects (17 of 22, 2 equal, $p < .05$) but nonsignificant defense effects (13 of 24, $p > .25$). These results could not be interpreted as contrary to the defense hypothesis since the requirement of an even distribution of responses to the stimuli in the base series was not met. The data were examined further, however, in order to determine the effect of the distinctive stimulus on the *differential ease of locating the critical and neutral pictures*.

Table 1 presents the distribution of mean frequency of response to each stimulus in the defense series. Condition A refers to the fact that all Ss were attempting to locate the position of the OS picture correctly, while under condition B, the MG picture was the relevant stimulus.

² The failure to alter the OE control picture did not appear to affect the obtained results.

TABLE 1

MEAN FREQUENCY OF RESPONSE TO THE FOUR STIMULI
IN THE DEFENSE SERIES OF EXPERIMENT I

	Condition A*			I
	OE	MG	OS	
Females	4.00	3.33	14.08	10.58
Males	2.33	3.34	16.58	9.75
	Condition B			
	OE	MG	OS	
Females	5.00	22.50	2.75	1.75
Males	5.83	21.42	3.25	1.50

* Condition A refers to the fact that all Ss were attempting to locate the OS stimulus; this was the critical picture for females and the neutral picture for males. In Condition B MG was the stimulus to be located; this was the neutral picture for females and the critical picture for males.

According to the defense hypothesis females would correctly locate OS (in condition A) less often than MG (in condition B). The prediction for male Ss was that they would locate the MG (in condition B) picture correctly relatively less often than the OS (in condition A) stimulus. However, the data in Table 1 indicate that both male (OS = neutral) and female (OS = critical) Ss have significantly fewer correct locations of the OS than of the MG stimulus ($t = 3.06$ and $t = 5.33$; $df = 11$; $p < .02$). Ten females conformed to the "defense" hypothesis as compared to only 3 male Ss.³ The primary determinant of errors in locating the "traumatic" picture, however, was a generalization tendency between the OS and the similar I stimulus. Thus, the unaltered I picture differentially affected the difficulty of identifying the position of the OS and MG stimuli. The negative results in the test of the defense hypothesis could be attributed to this artifact in the experiment. That is, the generalization tendencies among the stimuli may have obscured the operation of the defense process.

The question also arises, however, as to the extent to which the "perceptual defense"

³ A complete analysis of Blum's (2) results was not available, but examination of the data available indicates that a similar "sex difference" in the defense series occurred in that experiment as in the first study reported here. That is, the male Ss manifested very small "defense" effects mean correct response to both OS (neutral) and MG (critical stimulus) is approximately 24.00, whereas that of the female Ss is much larger (mean correct location of the OS, the *critical stimulus*, is 11.00; that of MG (neutral picture) was 21.14).

effects obtained by Blum (2) under similar experimental conditions are a function of this stimulus similarity variable rather than an active repression mechanism. The only control for stimulus properties was the essentially equal frequency of response to the four stimuli at .03 seconds exposure duration based on a preliminary group of Ss. The data from this experiment indicate the highly distinctive stimulus (at .03 exposure duration, I) *differentially* interfered with correct response to the experimental stimuli. This fact suggests that the stimulus properties which determine whether the stimulus "stands out" clearly at .03 seconds exposure duration is not identical (or equivalent) to those characteristics which determine the ease of correct discrimination at .20 seconds.

The following experiment was conducted in order to test the defense hypothesis under the specified conditions; i.e., an even distribution of responses to the stimuli in the base series. The selection of stimuli was such that it was also possible to explore the alternate hypothesis; i.e., differential error in locating the stimuli may be primarily a function of differential degrees of similarity between the critical stimulus and other stimuli in the set of four.

EXPERIMENT II

The I picture had proven to be perceptually distinctive in the base series and at the same time highly similar to the OS stimulus in the defense series. Therefore, an enlarged picture of Blacky from Card VII (A) was substituted and paired with the OS stimulus to form a critical-control pair for female Ss. The experimental critical-control pair for male Ss was formed by altering the OE picture until it was adjudged to be as similar to MG as A was to OS.⁴ The four pictures were then arranged on each of 8 cards so that the position of each picture was systematically rotated, every picture appearing twice in each position.⁵

Under these conditions the relative inability

⁴ Since Blum had obtained fairly equal frequency of response with MG, OS, and an altered OE stimulus these stimuli were retained. The selection of A to replace I was based on opinion of two judges that this picture was more similar to OS picture than to either of the other two. These judgments for similarity are, of course, not very precise; but were sufficient to serve the purposes of this experiment.

⁵ Blum (2) used only 6 cards.

of Ss to locate a critical stimulus as compared to a neutral one could be attributed to a perceptual defense mechanism. Failure to obtain these results, however, would lend support to the hypothesis that the "defense" effect noted in (2), as in the group of female Ss in Experiment I, was some function of differential variation in the degree of stimulus similarity among the test stimuli.

Method

Subjects and apparatus. Ten males and ten female Ss were randomly selected from a sophomore high school class. The apparatus consisted of a mirror tachistoscope with approximately .68 foot-lamberts illumination in each exposure field. Exposure duration in the test-field was controlled by a Hunter decade interval timer. When seated at the viewing window the S fixated a small dot in the center of the lighted pretest field which corresponded to the center of the four stimuli presented in the test field.

Procedure. The procedure and instructions were as similar as possible to those in the original experiment (2) with the following exceptions: (a) The sensitization instructions were typed and attached to the Blacky cards in order to minimize the interaction between E and Ss. (b) The changes in the control stimuli as indicated above.

Results. The choice of the stimulus (A) to replace the Identification picture and the alteration of the OE picture was such that one experimental and one control stimulus were judged to be more similar to each other than to either of the other two stimuli. The purpose of these changes was to reduce the distinctiveness of any one of the 4 stimuli at .03 seconds and, at the same time, to equalize generalization tendencies between each critical stimulus and one of the control pictures at .20 seconds. It was expected that the additional control of these stimulus properties would more nearly meet the conditions necessary for the test of the defense hypothesis and at the same time yield data relevant to the effect of stimulus similarity on the ease of discriminating the location of the stimuli in the defense series.

Examination of the data reveals that the change in stimuli did result in a more symmetrical distribution of responses by the groups in the base series. The respective means for the stimuli were: OE = 12.20; A = 12.55; MG = 16.50; and OS = 14.75. This distribution of scores would seem to meet the basic conditions necessary for testing the vigilance and defense hypotheses as outlined (2, 3).

Analysis of the data yielded significant vigilance effect when evaluated by the *t*-test ($t = 2.08$; $df = 19$; $.05 > p > .02$) although only 12 of 17 (3 equal) Ss conformed to the hypothesis. Only 9 Ss (of 20) conformed to the defense hypothesis. These results are not statistically significant at the .25 level of confidence as evaluated by the sign test. Analysis of the data with a more efficient statistical test also failed to yield a significant defense effect ($t = 1.60$; $df = 19$; $.20 > p > .10$).

TABLE 2
MEAN FREQUENCY OF RESPONSE TO THE FOUR STIMULI
IN THE DEFENSE SERIES OF EXPERIMENT II

	Condition A*			
	OE	MG	OS	A
Females	4.00	4.20	16.20	7.60
Males	4.00	3.50	17.00	7.50
	Condition B			
	OE	MG	OS	A
Females	8.00	15.70	4.30	4.00
Males	8.70	14.70	3.90	4.70

* Condition A refers to the fact that all Ss were attempting to locate the OS stimulus; this was the critical picture for females and the neutral picture for males. In Condition B MG was the stimulus to be located; this was the neutral picture for females and the critical picture for males.

The data relevant to the effects of the stimulus similarity variable on the ease of correctly identifying the position of the critical and neutral stimuli are presented in Table 2. The differences between the mean correct response to the OS and MG stimuli in no case (comparing males and females separately and as a group) reached the .20 level of confidence. However, statistically significant generalization effects were obtained within each critical-control pair. The mean for stimulus A, under condition A, is significantly larger than OE or MG ($.05 > p > .02$) while the mean of OE, under condition B, is significantly larger than OS or A ($.05 > p > .02$). These results demonstrate that the degree of similarity among the four stimuli is a significant determinant of the Ss' ability to identify correctly any one stimulus picture.

Discussion

The methodological implication of these results is quite clear. Variation in similarity among the four stimuli is an important determiner of the frequency of correct response in the defense series. Control of this variable is essential before any obtained differential response tendencies can be attributed to a perceptual defense mechanism. These results lend strong support to the hypothesis that a pre-experimental series of exposures at .03 seconds duration, even under conditions where the stimuli elicit essentially equal frequency of response, does not serve as an adequate control for the relevant stimulus properties. The fact that appropriate control for stimulus properties was not present in the original experiment (2) obviates the interpretation of these data in terms of a perceptual defense or repression mechanism.⁶ Whether or not sys-

⁶ Additional evidence supporting the repressed impulse interpretation of "perceptual defense" has

tematic errors of localization beyond those attributable to stimulus similarity are obtainable under these experimental conditions is a problem for future research.

There is some question however whether it would be necessary to postulate an active avoidance mechanism (repressed impulse and/or defense) in order to account for such results. Since the critical stimulus had been associated with anxiety, an alternative explanation in terms of the effects of anxiety on the generalization gradient (9) seems to be an equally plausible interpretation. Data from the replication experiment lend some tentative support to this hypothesis. The Ss who became sensitized to the critical stimulus (vigilance) tended to manifest a stronger generalization tendency between the critical-control pair of stimuli as compared to the neutral-control pair. Further experimentation is under way explicitly to test this hypothesis.

The data from the two experiments reported here do offer support for the vigilance hypothesis. In the first experiment, 17 (of 22, 2 equal) Ss were more responsive to the critical stimulus in the vigilance than in the base series ($p = .05$). The results of the second experiment were not statistically significant according to the sign test; but a t test yielded a ratio of 2.08, which for 19 degrees of freedom is significant between the .05 and .02 level of confidence. The combined probability of both experiments is significant beyond the .05 level of confidence ($\chi^2 = 10.84$; $df = 4$). These results are consistent with other experimental evidence indicating that association of anxiety with previously neutral stimuli (7, 8) or anxiety arousal under certain experimental conditions (6, 12, 13) is related to an increased perceptual sensitivity to anxiety cues.

It should be pointed out however that the increased responsiveness to the critical stimulus does not necessarily involve any "gratification of repressed impulses." For example, it has been demonstrated that perceptual sensitization tends to occur if identification of the anxiety cue prevents further anxiety arousal (8) and under experimental conditions which tend to "accentuate" the anxiety producing

been recently reported (1). The similarity between the criterion for repression as a characteristic defense mechanism and the method of testing the perceptual consequences of repression, however, make interpretation of the data equivocal.

stimulus (7). The anxiety-producing instructions accompanying the critical stimulus, as compared to the neutralizing instructions associated with the other stimulus, may have served to produce such accentuation. The increased responsiveness could be attributed to the effects of differential learning resulting from the experimental sensitization experience, i.e., to the differential familiarity of the critical stimulus (4, 10, 11).

SUMMARY

The results of an initial attempt to replicate the results of Blum's experiment (2) indicated that stimulus similarity was an important determinant of errors of localization in the test of the perceptual defense hypothesis. A second experiment was designed to test the defense hypothesis under the specified conditions and at the same time yield information concerning the role of stimulus similarity on errors of recognition in the defense series. The results indicated similarity among the experimental and control stimuli was the primary determinant of the frequency of correct response. Therefore, it was concluded that relatively equal frequency of response to these stimuli at .03 seconds exposure duration is not an adequate control for the effects of stimulus properties on errors of localization at .20 seconds exposure duration. Further, since appropriate control of stimulus factors was not present in the original experiment (2), serious doubt arises concerning the perceptual defense interpretation of results obtained in that experiment. Evidence was presented which tentatively suggests that systematic errors of localization might be due to an increased generalization gradient associated with anxiety arousal.

The data from both experiments reported here confirmed the findings with respect to the increased perceptual sensitivity to stimuli associated with anxiety (perceptual vigilance). It was pointed out, however, there is no necessary reason to attribute these results to repressed impulses. An alternative explanation in terms of possible differential learning under conditions of anxiety arousal was suggested.

REFERENCES

1. BLUM, G. S. Perceptual defense revisited. *J. abnorm. soc. Psychol.*, 1955, **51**, 24-29.

2. BLUM, G. S. An experimental reunion of psychoanalytic theory with perceptual vigilance and defense. *J. abnorm. soc. Psychol.*, 1954, **49**, 94-98.
3. BLUM, G. S. ADI Document #4031. Library of Congress.
4. HOWES, D. H. On the interpretation of word frequency as a variable affecting speed of recognition. *J. exp. Psychol.*, 1954, **48**, 106-112.
5. HOWES, D. H., & SOLOMON, R. L. Visual duration thresholds as a function of word probability. *J. exp. Psychol.*, 1951, **41**, 401-410.
6. LINDNER, HAROLD. Perceptual sensitization to sexual phenomena in the chronic physically handicapped. *J. clin. Psychol.*, 1953, **9**, 67-68.
7. LYSAK, W. The effects of punishment upon syllable recognition thresholds. *J. exp. Psychol.*, 1954, **47**, 343-350.
8. REECE, MICHAEL M. The effect of shock on recognition thresholds. *J. abnorm. soc. Psychol.*, 1954, **49**, 165-172.
9. ROSENBAUM, G. Stimulus generalization as a function of the level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35-43.
10. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256-270.
11. SOLOMON, R. L., & POSTMAN, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43**, 195-201.
12. SPENCE, D. P. Vigilance and defense as anxiety manifestations. *Amer. Psychologist*, 1955, **10**, 387. (Abstract)
13. STEIN, K. B. Perceptual defense and perceptual sensitization under neutral and involved conditions. *J. Pers.*, 1953, **21**, 467-478.

Received August 25, 1955.



POLITICAL AND CHILD-REARING ATTITUDES IN SWEDEN

RICHARD H. WILLIS

University of Wisconsin

THE problem investigated in the present study is the relation between the acceptance of a role of authority and the acceptance of a role of submission to authority. It is hypothesized that a tendency to use freely the authority of the former role is associated with a tendency to submit freely to the authority encountered in the latter role. Substantiation of this hypothesis would indicate the existence of a personality variable which may be termed *admiration of authority*. Persons possessing a high degree of admiration for authority, by the present hypothesis, seek out positions of authority and freely exercise their authority over subordinates, but at the same time submit readily to authority in the hands of others. The pathological analogue of this type of personality structure has been perceptively analyzed by Fromm (3) in terms of sadism and masochism, while Adorno *et al.* (1) in their study of prejudice discuss six personality syndromes revealed by questionnaire and interview data, one of which, the authoritarian type, has as one of its defining characteristics an identification with strength, i.e., an admiration of authority.

The role of *parent* was taken to represent the role of authority and that of *citizen* to represent that of submission to authority. The hypothesis of primary interest was then formulated as:

Hypothesis 1. *Individuals who demand (or who would demand) strict obedience from their children tend to accept demands for strict obedience made on them as citizens.*

Two groups of items were constructed to test this primary hypothesis. The first of these was the POL group (Political—acceptance of authoritarian political policies), while the second item group was the DFO (Demands for Obedience) group. Two other item groups were constructed in order to explore two other areas judged to be relevant. The UPB (up-bringing) group was constructed from items concerning relatively factual aspects of the individual's upbringing.¹ By noting the correl-

ation between POL and UPB, the following hypothesis is tested:

Hypothesis 2. *Individuals who, as children, accept demands for strict obedience from their parents tend to accept demands for strict obedience made upon them as citizens.*

Then by means of the correlation between DFO and UPB, a third hypothesis may be tested:

Hypothesis 3. *Individuals who, as children, accept strict demands for obedience from parents tend to demand strict obedience from their children.*

These three hypotheses are central in the present study. However, as background data on respondents were also available, it was possible to test additional hypotheses regarding differences in subgroup means. The five variables used for this purpose were sex, age, occupational status, father's occupational status, and educational level.

METHOD

The present study is an analysis of part of the data collected in a more comprehensive survey of political attitudes, the Uppsala Survey,² conducted by the Sociology Institute of the University of Uppsala, Sweden, during the winter of 1951-52. The city of Uppsala has a population of about 60,000. As well as possessing the country's oldest university, which has an enrollment of approximately 4,000, it is situated in an agriculturally fertile region, and also has a certain amount of industry.

A simple random sample of adults (18 and over) was obtained, using the files of the city registrar's office. Approximately 220 individuals were selected and from among these 194 were reached and interviewed. Thirty-four interviewers were used, all of whom were university students studying sociology. Of these, 32 interviewed 6 persons each, and the remaining 2 interviewers contacted 1 person each. Each interviewer had received training in interviewing techniques and had passed an achievement test in this area. Names of persons to be interviewed were assigned at random.

it made no contribution to the interpretation of the results, as a chi-square test indicated that it did not discriminate between subgroups of the sample. This lack of discriminative power can probably be attributed to the relatively strong wording of the items, which made them unacceptable to all but a small proportion of the respondents, and resulted in a highly skewed distribution of responses.

² Under the direction of Hans L. Zetterberg.

¹ A fourth item-group was constructed, which was intended to tap the respondent's feelings of rebelliousness toward his parents and his home environment, but

Of the 194 persons interviewed, 144, or 74 per cent, were used in the study. There were 73 men and 71 women. The remainder represent cases in which certain information was missing from the interview schedule. In almost all cases the missing data were responses to some of the items in the three item-groups POL, DFO, and UPB. In a very few cases the missing data involved background information on the interviewee—sex, age, etc. When more than two items in any item-group were unanswered, the interview schedule was not included in the tabulation. For those interview schedules in which not more than two items were unanswered in any item-group, an interpolative scoring procedure was used, the unanswered items being assigned intermediate scores. For the most part, missing responses probably represent inability of the interviewee to make up his mind about an item. Interviewers were instructed to use some persuasion in eliciting responses, but not to insist. A few interviewees may have been unwilling to commit themselves on certain topics even though they were in all cases informed that their answers were to be used only for research purposes.

The items were initially worded in English by the author and then translated into Swedish by a Swedish sociologist who rephrased items, when appropriate, to fit the national and local verbal habits. The items were then presented for comments to the members of the methodology class from which the interviewers were drawn and some items were reworded on the basis of the suggestions received. All items were then pre-tested on a small group of adults. The main change resulting from this pretest was to reduce the number of choices of answers to some questions from five to three so that all items were either dichotomous or trichotomous in their final form.

The POL group contains four items concerning freedom of political expression, two about national honor, two more dealing with unquestioning obedience to the state, one about punishment, and a remaining item pertaining to restriction of group activities. The major subgroup within DFO consists of seven items about punishment of children. Another subgroup of four items deals with obedience of children to decisions of parents, while the remaining two items maintain that parents know best. The items in UPB all inquire about relatively factual matters of the respondent's upbringing. Six items relate to freedom of activities as a child, while six others inquire about the respondent's voice as a child in child-parent relations. Thus, POL contains ten items, DFO thirteen, and UPB twelve. High scores signify a totalitarian attitude (POL), authoritarian child-rearing attitudes (DFO), or an authoritarian upbringing (UPB).³

RESULTS AND DISCUSSION

The means, standard deviations, and test-retest reliabilities of the three item-groups are

³ The items are not given here, but have been deposited with the American Documentation Institute. Order Document No. 4838 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1
SCORE DISTRIBUTIONS AND RELIABILITY OF
ITEM GROUPS

Item-group	Highest Possible Score	Mean	SD	Test-retest Reliability
POL	20	11.62	3.91	.73
DFO	38	16.72	7.20	.68
UPB	24	8.76	4.88	.83

TABLE 2
ITEM-GROUP INTERCORRELATIONS

Item-groups	Correlation Coefficient	
	Uncorrected	Corrected for Attenuation
POL-DFO	.33*	.43
POL-UPB	.22*	.27
DFO-UPB	.21*	.28

* $p < .01$.

shown in Table 1. Also shown is the highest possible score for each group. All groups exhibit means well removed from either end of the range of possible scores and the standard deviations indicate satisfactory spread of scores. To obtain an estimate of the item-group reliabilities, 36 persons were reinterviewed at a later date with the same interview schedule. The obtained reliability coefficients are quite satisfactory for short instruments designed to detect group effects only.

Table 2 shows the correlations between item-groups, both as computed directly and as corrected for attenuation. The correlation of primary interest is that between POL and DFO, as it is this which indicates the validity of Hypothesis 1. It is the largest of the correlations, being .33 uncorrected and .43 corrected for attenuation. A two-tail z -transformation test of significance shows this correlation to be significant beyond the .01 level of confidence. A correlation ratio was also computed from a 9 by 10 scatter diagram which, when corrected for discontinuity, was equal to .36. A chi-square test of the significance of curvilinearity (4) showed the p value to be greater than .40 and therefore the hypothesis of linearity of regression is not rejected.

POL-DFO correlations within subgroups (sex, age, occupational status, father's occupational status, and educational level) ranged from .15 to .72 but two-tail z -transformation

tests of significance indicated that no difference between subgroup correlations were significant at the .05 level.

Partialling out the effects of UPB yields a POL-DFO correlation of .30, indicating that upbringing as measured by UPB has little connection with political views outside that resulting from its communality with DFO. This finding is verified by the slight improvement in the prediction of POL scores resulting from using both DFO and UPB scores, the multiple r being .37.

Turning to Hypothesis 2, it is supported by an uncorrected r of .22 between POL and UPB, significant at the .01 level. Hypothesis 3 is also accepted on the basis of the DFO-UPB correlation of .21, also significant at the .01 level. Thus all three of the major hypotheses are accepted at the .01 level of confidence.

Two additional partial correlations were computed. The first shows that the association between POL and UPB is lowered from .22 to .17 when DFO is partialled out. Likewise, the association between DFO and UPB is lowered from .21 to .14 by partialling out the effects of POL. This shrinkage of correlations is interpreted as signifying that Hypotheses 2 and 3 are verified in part as a consequence of the POL-DFO correlation rather than as entirely independent relationships.

Although all three intercorrelations between item-groups are significant, it is important to note that the magnitudes of the r 's are such that only a small fraction of the total variance of scores on one item-group is associated with the variance of the scores on either of the other two item-groups. Would an American sample have exhibited higher intercorrelations? Most likely not. More recently two groups of items

which were quite similar to, and in some cases identical with, the items in POL and DFO were administered to a group of 100 undergraduates at the University of Wisconsin. Interestingly enough, the correlation obtained was .33—exactly that found between POL and DFO. It seems reasonable to assume that an American sample would yield about the same magnitude of correlations between POL and UPB, and DFO and UPB as did the Swedish sample.

For any sample, it seems likely that the intercorrelations would have been higher if the roles involved had been perceived by the respondents as belonging to a single hierarchy of authority. For example, every Army officer must take both the role of a superior (toward his subordinates) and of a subordinate (toward his superiors), and the two roles are clearly perceived as fitting into a single chain of authority. Under these conditions, it would be expected that free use of authority, i.e., strict demands upon subordinates, would be more closely associated with ready compliance to authority from above than in the case of the roles investigated in the present study.

Some interesting differences in subgroup means were obtained when the sample was divided according to certain antecedent factors (Table 3). There were three categories for each factor other than sex. By far the most highly significant difference in means is that between levels of educational attainment on POL. The F ratio required to reach the .01 level of significance, with 2 and 141 degrees of freedom, is 4.76 while that obtained was 21.63. The more school grades completed, the less political authoritarianism was shown. The more educated groups also had significantly lower scores on DFO and UPB. Age too was influential, the older groups being significantly higher on all three item groups. A peculiarity of the occupational status factor was that the means were nonmonotone, the middle group being the lowest on POL and DFO, but highest on UPB. In contrast, those whose father's occupational status was relatively high were the lowest on POL and DFO, while the highest UPB mean was for the intermediate level of the father's occupational status. The only significant sex difference was that men were lower on POL than were women.

It is of importance that these antecedent

TABLE 3
LEVELS OF SIGNIFICANCE FOR DIFFERENCES IN
SUBGROUP MEANS

	Item-group		
	POL	DFO	UPB
Sex	.01	—	—
Age	.01	.05	.01
Occupational status	—*	.05*	—*
Father's occupational status	.01	—	.05*
Educational attainment	.001	.05	.05

* Nonmonotone means.

factors affect the item-group means appreciably, but have no significant effect on the correlation between POL and DFO, as mentioned earlier. This finding supports the view of a dynamic relationship between roles of dominance and those of submission, such as is taken by Fromm in his analysis, but does not disclose the specific nature of the relationship.

SUMMARY

Three groups of questionnaire items were constructed to measure authoritarian political attitudes, authoritarian child-rearing attitudes, and authoritarian aspects of upbringing. One hundred forty-four adults from a random sample in a Swedish town were interviewed. It was found that individuals who made strict demands as parents expected in turn that strict demands be made on them as citizens. Significant positive correlations were also

found between authoritarianism of upbringing and authoritarianism of both political and child-rearing attitudes. These findings verified predictions which had been derived from three hypotheses concerning the nature of the relationship between roles of authority and roles of submission.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ERIKSON, E. H. Hitler imagery and German youth. In C. Kluckhohn, & H. A. Murray, (Eds), *Personality in nature, society, and culture*. New York: Knopf, 1950.
3. FROMM, ERICH. *Escape from freedom*. New York: Rinehart, 1941.
4. GARRETT, H. E. *Statistics in psychology and education*. (3rd Ed.) New York: Longmans, Green, 1947.

Received September 8, 1955.

THE NEGRO STEREOTYPE AND PERCEPTUAL ACCENTUATION¹

PAUL F. SECORD,² WILLIAM BEVAN, AND BRENDA KATZ³

Emory University

THE tendency on the part of individuals in our culture to stereotype the Negro by attributing to him certain unfavorable personality characteristics has been studied by social psychologists for several decades. Viewing this useful and familiar concept of the stereotype from the perspective of contemporary work in social perception, however, raises several questions that have not been directly investigated.

First, is the impression of personality traits obtained from a face controlled by such categorical responses as, "This is a white man," or "This is a Negro"? Are light Negroes and dark Negroes stereotyped to the same degree? Or, is there, instead, a quantitative relationship between the degree to which a Negro possesses "Negroid" physical characteristics and the extent to which stereotyped traits are attributed to him? Such a categorical response—assigning a set of traits to an individual because of his membership—has generally been a part of the definition of a stereotype, although empirical data have seldom been presented to support this view in any direct fashion.

Secondly, do anti-Negro individuals exaggerate stereotyped personality characteristics in rating Negro photographs, and do pro-Negro individuals de-emphasize these characteristics, as compared with neutral judges? An affirmative answer is anticipated. Prejudiced persons may be expected to rate Negro faces lower on favorable traits and higher on unfavorable traits; the reverse is likely to be true for pro-Negro persons.

Finally, do anti-Negro and pro-Negro individuals perceive the Negro as "more Negroid" in physical traits than an unprejudiced judge? This question involves an extension of Bruner and Postman's (7) concept of accentuation, which has generally been described in operational terms as an increase in perceived and remembered size occurring when stimuli are valued or need-related. In the present paper

the concept is extended beyond mere size to ratings of physiognomic variables used to characterize Negroidness as compared with Caucasian-like qualities. Although such variables, unlike "size," cannot be as readily expressed in physical terms, they are dimensions which might be expected to reveal accentuation, because of their close relation to a personal variable such as prejudice.

METHOD

Photographs. Fifteen Negro pictures, ranging from very light to very dark, were chosen from yearbooks of two Negro colleges in Atlanta, Georgia. These pictures were rated by a preliminary group of student judges on 15 physiognomic traits believed to be characteristic of Negroes. Ten photographs were retained for further use. Five white photographs were also chosen from yearbooks of two Southern universities. In addition to the attempt to obtain photographs which roughly ranged along a continuum of Negroidness, several other criteria were used for selecting pictures. All were full face, black and white photographs, approximately the same size (about 2" × 2"), of young adult males. None had distinctive or unusual appearances; e.g., photographs were omitted if the individual wore glasses, had a mustache, or unusual facial marks. Pictures of individuals who were definitely smiling, frowning, or exhibiting other strong signs of emotion were not included. Portions of the photographs showing clothing were removed.

Subjects. Initially, 47 students, 15–18 years of age, were obtained from high schools in Atlanta, Georgia. A Likert-type scale of prejudice to be described later was administered to them. Upon tabulating their prejudice scores, very few of them were found to have scores at the end of the scale involving "favorableness toward the Negro;" consequently the group was augmented by additional subjects from a less prejudiced population.

An attempt was made to obtain these additional students from school systems in New York and New Jersey; however, school administrators refused permission to conduct the study because it involved Negro photographs. The extra subjects finally chosen were from a B'nai B'rith group in Atlanta. This is a Jewish organization which stresses tolerance and brotherhood, and thus is not strictly matched with the "anti-Negro" and the "neutral" groups of judges. Five girls and six boys from this group, 15–17 years of age, volunteered to participate.

Rating scales. Two rating scales were developed for this experiment. In order to insure that trait-names would be expressed in language familiar to the rater, 16 volunteer students were asked to list the physical and personality traits that they, or people in general, usually associate with the Negro. The final choice of the 10

¹ This is the eighth in a series of papers entitled, "Personalities in Faces."

² Now at the University of Nevada.

³ Now at Washington University.

physiognomic traits used in the rating scale took into account these preliminary descriptions as well as physical characteristics of Negroes described in several anthropological references (3, 13, 15). Physiognomic traits were rated on a seven-point numerical scale, the ends of the scale being designated by appropriate adjectives. For example, the trait *width of nose* was labeled *wide* at the 1 end and *thin* at the 7 end.

The 7-point personality rating scale consisted of 15 traits chosen from the preliminary descriptions and from the traits given the Negro by the subjects in Gilbert's repetition (9) of the early Katz and Braly study (11) of racial stereotypes. Certain traits believed to be obviously unsuitable for rating from a facial photograph, e.g., "ostentatious," were not included in the scale. Personality traits were not given adjectival labels at the ends of the scale, but were labeled *most characteristic* at either the 1 end or 7 end, and *least characteristic* at the opposite end. After the rating procedure, judges in the B'nai B'rith group were shown the pictures again and asked to identify which were Negro and which were White. This precaution was not exercised with the earlier groups.

Negro attitude scale. No existing scale of attitude toward the Negro appeared to be entirely satisfactory. Two common faults were that some items were dated by their content, and others were too extreme to discriminate adequately varying degrees of prejudice. Consequently a composite scale of 18 items was compiled from the attitude scales in Adorno, *et al.* (1), Likert (14), and Hinckley (10).⁴ These were administered and scored by utilizing the Likert method of five response alternatives for each item.

RESULTS

Obtained ratings on the physiognomic scale were transformed directionally so that in each case a *low rating* implied more Negroidness, and a high rating more Caucasian-like appearance. Ratings by each judge on the 10 physiognomic traits were then added to obtain a total Negroidness score for each photograph rated by him. As a test of the appropriateness of this procedure, each of the 10 items was correlated with the total Negroidness score. As may be seen from Table 1, eight of them correlated very highly with the total score. Indeed, the correlation of .97 between dark-light complexion and Negroidness suggests that the latter can be rated very accurately on the basis of a single-item scale. In data presented in the remainder of the paper, however, physiognomic Negroidness is based on a total score for the eight items, omitting *cheekbones* and *curve of mouth*. This score is abbreviated *Neg*.

Next, the ratings for the 15 personality at-

tributes were transformed directionally so that a *low rating* was in the direction of the Negro stereotype. On every trait, the mean rating for the white photographs was compared with the mean for the Negro photographs, using *t* tests. The prediction that Negroes would be rated *lower* than whites on the following traits, alert expression, honest face, air of responsibility, air of refinement, intelligent look, and thrifty, was confirmed in all instances ($p < .05$). Again in accordance with expectations, Negroes were rated higher than whites on the following: superstitious, lazy, emotional, untidy, and immoral. Four other traits on which the Negro was expected to be rated higher were: happy-go-lucky and pleasure-loving, on which there was no significant difference ($p > .05$), and warmhearted and religious, which were rated in the opposite direction. The 11 traits rated in the prediction direction ($p < .05$) were combined to obtain a total score for personality stereotyping (abbreviated *Ster*).

Our first hypothesis states that there is no difference in the degree to which Negro photographs varying in physiognomic Negroidness (*Neg*) yield stereotyped personality impressions. That is, even if he has a Caucasian-like appearance, a Negro will be seen as having, in full degree, all the stereotyped traits usually attributed to the Negro. In order to test this hypothesis, the mean *Neg* score was obtained for all judges on each photograph. The mean *Ster* score was also obtained for each picture. These data are presented in Fig. 1.

From the curve for the mean *Neg* scores, it may be seen that the 10 Negro photographs

TABLE 1
CORRELATIONS BETWEEN EACH PHYSIOGNOMIC TRAIT
AND THE TOTAL PHYSIOGNOMIC RATING FOR
EACH PICTURE

Physiognomic Traits	r^*
Complexion (dark)	.97
Hair (coarse)	.94
Fullness of lips (thick lips)	.90
Skin texture (oily)	.89
Nose (flat)	.89
Nose (wide)	.88
Hair (curly)	.85
Ears (small)	.81
Curve of mouth (corners turned down)	.55
Cheek-bones (prominent)	.24

* $r = .64$ at the .01 level, with $N = 15$.

⁴ The writers are indebted to Dr. Hinckley for furnishing his scale items in a personal communication.

and the five white pictures become increasingly Caucasian, from left to right in the figure, and that the Negro photographs are continuous with the white, on this variable. The *Ster* score curve, on the other hand, does not at all follow the physiognomic one, but instead, has essentially a horizontal slope. Thus, personality stereotyping remains unchanged with increasing Caucasian appearance. In fact, data for pictures 9 and 10 indicate that the two whitest Negroes are as heavily stereotyped as the darkest Negroes, in spite of the probability that, if the photograph of the lightest Negro had been shown in a context of white photo-

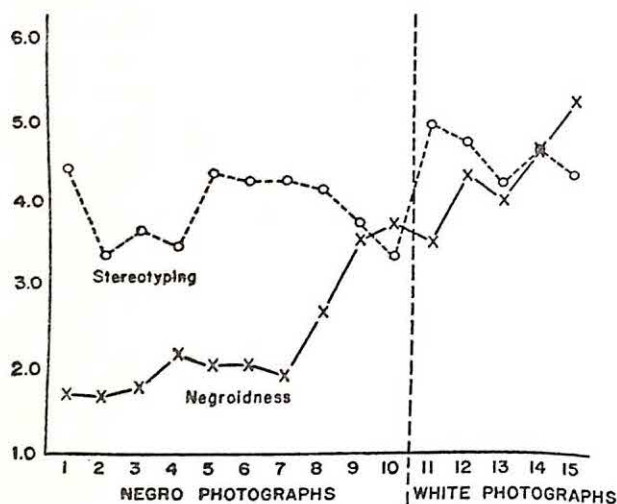


FIG. 1. A COMPARISON OF THE MEAN PHYSIOGNOMIC NEGROIDNESS SCORE FOR EACH PHOTOGRAPH WITH THE CORRESPONDING INDEX OF PERSONALITY STEREOTYPING

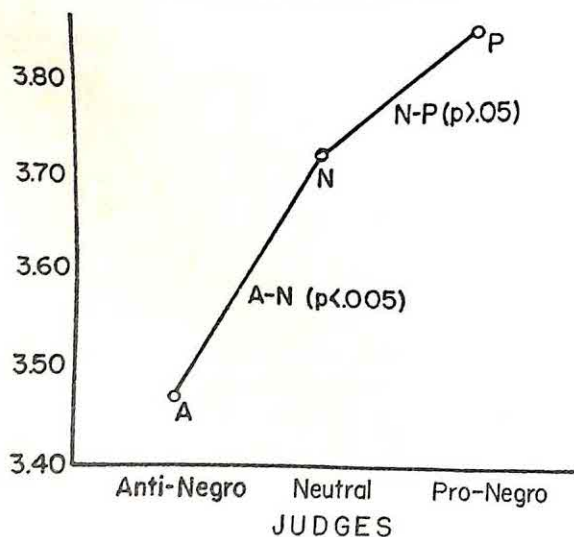


FIG. 2. MEAN PERSONALITY STEREOTYPING SCORES FOR ANTI-NEGRO, NEUTRAL, AND PRO-NEGRO GROUPS

graphs only, he most likely would not have been recognized as a Negro.

The conclusion to be drawn from Fig. 1, then, is that, in the eyes of our high school subjects: "A 'Nigger' is a 'Nigger,' no matter how white he may be." It should be noted that, according to their own statements, the subjects had no difficulty in identifying nine of the Negro photographs as Negroes, and only slight difficulty was reported in the case of the lightest Negro. Apparently, identification of the photograph as Negro is sufficient to yield the full Negro stereotype, regardless of how Negroid or how Caucasian the individual looks. Because of the possibility that prejudiced judges might show this effect to a greater extent than neutral ones, the 58 judges were divided into groups on the basis of their prejudice scores, as follows: "anti-Negro," 1.0-2.4; "neutral," 2.5-3.4; and "pro-Negro," 3.5-5.0. Separate curves for the three groups of judges were found to be very similar to Fig. 1, indicating that all three groups respond categorically to Negro photographs.

Our second hypothesis was that anti-Negro judges would exaggerate the personality stereotype of the Negroes, whereas the pro-Negro group would de-emphasize it. To test this hypothesis, the mean *Ster* score on the 10 Negro photographs was determined for each group of judges. The location of these mean scores is in accord with the hypothesis, as Fig. 2 shows. Thus, the anti-Negro group intensifies the stereotyped characteristics of the Negroes, as compared with the neutral group. The difference between the neutral group and the pro-Negro group, however, is not sufficiently large to reach significance at the .05 level.

Another way of testing the hypothesis is to correlate the Likert prejudice score for each judge with his *Ster* score for the Negro photographs. An r of .31 ($df = 56$, $p = .01$) was obtained. Although this finding supports the hypothesis, its low magnitude suggests that self-ratings of prejudice and ratings of the personality characteristics from Negro photographs are operations having quite different psychological significance.

Finally, the less obvious hypothesis was tested that both anti-Negro and pro-Negro judges would perceive the Negro as more Negroid in physiognomic traits than would

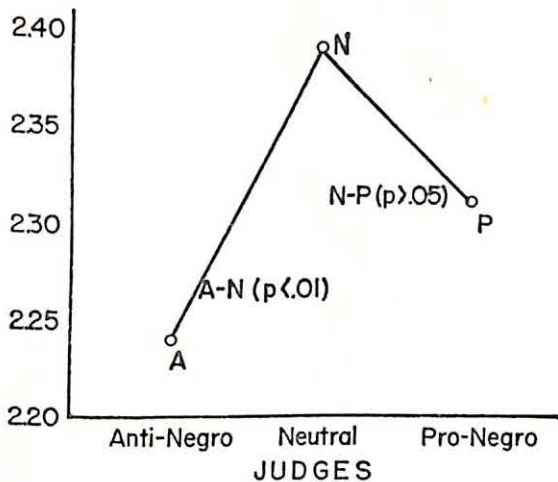


FIG. 3. MEAN INDICES OF PHYSIOGNOMIC NEGROIDNESS FOR ANTI-NEGRO, NEUTRAL, AND PRO-NEGRO GROUPS

neutral judges, by a process of accentuation. From Fig. 3, it is clear that this hypothesis is upheld for the anti-Negro group, but not for the pro-Negro judges. The location of the mean *Neg* scores for the three groups of judges agrees with the hypothesis; however, only the difference between the anti-Negro and neutral group is significant at the accepted level. A further test of this hypothesis was made by computing a correlation ratio between the Likert prejudice score for each judge and his *Neg* mean, but this was not significant ($\eta = .14$, $df = 56$, $p > .05$).

DISCUSSION

Although responding to a member of a minority group categorically—assigning a set of traits to an individual because of his group membership—has generally been a part of the definition of a stereotype, empirical data have seldom been presented to support this notion in any direct fashion. Instead, it is usually implicitly assumed that if an individual exhibits prejudice toward Negroes on an attitude scale, he will express those feelings toward persons specified as Negroes, regardless of their characteristics. There is recognition, of course, that a Negro in a special face-to-face relationship to an individual may be differently perceived; e.g., one's own Negro "Mammy," but little attention has been paid to more general dimensions along which Negroes may differ and the perceptual variations they may produce. Our data have been specifically concerned with the distinguishing physical characteristics

of the Negro, and we have shown that in perceiving photographs which vary widely in these characteristics, judges with varying degrees of prejudice are alike in that they all respond categorically. This evidence constitutes empirical support for the classical concept of the stereotype as a categorical response.

The low but significant relationship between prejudice scores and degree of stereotyping deserves further comment. Although both may be regarded as measures of prejudice, the operations involved are apparently sufficiently different to give rather divergent scores. The following are suggested as being among those factors differentiating the two operations: (a) Endorsement of an anti-Negro scale item identifies the rater as prejudiced, whereas assignment of an unfavorable trait to a photograph is not clearly identified with prejudice; (b) In rating photographs, physiognomic cues other than racial determine, to some degree, the attribution of personality traits; (c) A prejudice scale involves a conscious and deliberate effort on the part of the subject to indicate his attitude toward Negroes, whereas rating personality traits from faces appears to be an intuitive, effortless procedure.

There is an obvious difference in content, of course, between rating faces on such traits as *intelligent look*, *superstitious*, etc., on the one hand, and stating whether or not Negroes should be segregated in schools, buses, churches, hospitals, recreational facilities, and places of employment, on the other. Some attitude items, however, concern traits or qualities of the Negro and, therefore, are not appreciably different in content from trait ratings of faces, although as indicated above, the frame of reference within which ratings are made may be quite different for the two operations. The use of ratings of Negro photographs as a means of assessing the degree of prejudice perhaps deserves further attention as an alternative to the more familiar attitudinal techniques of measurement.

Laboratory experiments on accentuation have concerned themselves primarily, although not exclusively, with size. There is no reason why more complex social variables might not be studied in the same fashion. In fact Bruner (4) has suggested that preference

should be given to variables involving cues that are relevant to personality factors.

Accentuation may be regarded in general terms as a process of perceiving valued objects in a manner which enhances their value. Value is enhanced through distortion of those dimensions of the object which are value-relevant. Thus, in the case of perceiving coins as slightly larger, size is thought to be value-relevant because larger coins, generally speaking, are worth more. In the case of symbols, such as the swastika, which has been used by Bruner and Postman (6) and Klein, Schlesinger, and Meister (12), size is not directly value-relevant—there is no inherent reason why a larger swastika should have more or less value than a smaller one. As Bruner (4) has noted, however, increment in value might be derived from the fact that, in our culture, higher value is generally attached to larger things.

Taken as a whole, however, the empirical findings on size-accentuation are by no means entirely consistent or clear (cf. Allport, 2). In particular, the rationale for the relevance of size to value is of doubtful validity. The dimensions of physiognomic Negroidness in the present study, on the other hand, are clearly relevant, because the Negro is defined in physiognomic terms, and our value-measure is attitude toward *Negroes*. Negroidness is a negative value, and as such is accentuated or made more pronounced by our prejudiced judges.

Some comment on the performance of the pro-Negro group is in order. In a number of previous experiments on size-accentuation (5, 6, 8, 12), both positively and negatively valued objects have sometimes been accentuated. From these studies it might be predicted that the pro-Negro group would likewise accentuate Negroid features. Our interpretation of this empirical prediction in terms of the dynamics involved is that, although pro-Negro subjects may have adopted "tolerance for minority groups" as an important value, this value has not been fully integrated into their basic personality structure. These subjects might thus be expected to "lean over backwards" to avoid the appearance of prejudice, and hence are oversensitive to the identifying characteristics of a minority group.

Although the pro-Negro group rated the photographs more Negroid than did the neutral

group, this difference was not statistically significant. In this regard, a difficulty in the present design is the arbitrariness of identifying the "neutral" point of the prejudice scale. Conceivably, some of the pro-Negro judges were "really neutral," so that the difference between the neutral and the pro-Negro groups as chosen was attenuated. The small size and the somewhat atypical nature of our pro-Negro sample prevent us from drawing certain conclusions.

SUMMARY AND CONCLUSIONS

A series of 15 photographs, 10 of Negro faces, and 5 of White faces, were rated by 58 judges on two 7-point scales. One of these scales consisted of 10 physiognomic traits known to be characteristic of the Negro race, and the other, of 15 personality characteristics generally accepted as comprising the Negro stereotype. Ratings on the Negro and the White photographs were next compared for each trait. Those traits which failed to discriminate between the two sets of photographs were eliminated. For every photograph, two basic scores were obtained from each scale by adding together the scores for all judges on the remaining traits. Thus, a score for *physiognomic Negroidness* and for *personality stereotyping* was obtained. Negro photographs had been previously selected to insure a considerable range of variation in Negroidness. Prejudice scores were also obtained by administering a Likert-type scale of attitude toward the Negro.

The hypotheses tested and the findings obtained are as follows:

The first hypothesis was that there is no difference in the degree of personality stereotyping of Negro photographs varying widely in physiognomic Negroidness. That is, even if he has a Caucasian-like appearance, a Negro will be seen as having, in full degree, all the stereotyped traits usually attributed to the Negro. Supporting evidence was obtained: there was no decrease in stereotyping in moving from the most Negroid Negroes to the most Caucasian. The generally accepted but seldom tested definition of a stereotype as a categorical response to a member of a minority group is thus upheld.

The second hypothesis was that anti-Negro judges exaggerate the personality stereotype

of Negroes, whereas pro-Negro judges de-emphasize it. Comparison of the mean stereotype ratings of judges put into groups on the basis of their prejudice scores (anti-Negro, neutral, and pro-Negro) lent significant support to the hypothesis. Apparently, indicating degrees of agreement with pro-Negro or anti-Negro statements, as on the Likert scale, has something in common with rating Negro faces on valued traits—both may be regarded as ratings of prejudice.

Finally, the hypothesis of perceptual accentuation was tested for physiognomic traits: both anti-Negro and pro-Negro judges were expected to perceive the Negro as more Negroid in physiognomic traits than do neutral judges. Comparison of the mean Negroidness scores for the three groups of judges supported the hypothesis, although the difference between the neutral group and the pro-Negro group was not statistically significant. This partial lack of support for the hypothesis might be due to the atypical nature of the pro-Negro sample, which was drawn from a different population than the other two groups, or from a difficulty in defining what is "neutral" and what is "pro-Negro" on a Likert-type scale.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, F. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
3. ASHLEY-MONTAGU, M. F. *An introduction to physical anthropology*. (2nd Ed.) Springfield, Ill.: Charles C Thomas, 1951.
4. BRUNER, J. S. Personality dynamics and the process of perceiving. In R. R. Blake, & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 121-147.
5. BRUNER, J. S., & GOODMAN, C. D. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
6. BRUNER, J. A., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. soc. Psychol.*, 1948, **27**, 203-208.
7. BRUNER, J. S., & POSTMAN, L. Perception, cognition and behavior. *J. Pers.*, 1949, **18**, 14-31.
8. DUKES, W. F., & BEVAN, W., JR. Size estimation and monetary value: a correlation. *J. Psychol.*, 1952, **34**, 43-53.
9. GILBERT, G. M. Stereotype persistence and changes among college students. *J. abnorm. soc. Psychol.*, 1951, **46**, 245-254.
10. HINCKLEY, E. D. The influence of individual opinion on construction of an attitude scale. *J. soc. Psychol.*, 1932, **3**, 283-296.
11. KATZ, D., & BRALY, K. W. Racial prejudice and racial stereotypes. *J. abnorm. soc. Psychol.*, 1935, **30**, 175-193.
12. KLEIN, G. S., SCHLESINGER, H. J., & MEISTER, D. The effect of personal values on perception: An experimental critique. *Psychol. Rev.*, 1951, **58**, 96-112.
13. LEWIS, J. H. *The biology of the Negro*. Chicago: Univer. of Chicago Press, 1942.
14. LIKERT, R. A technique for the measurement of attitude. *Arch. Psychol.*, N. Y., 1932, No. 140.
15. STIBBS, E. P. *Physical anthropology*. London: Edward Arnold, 1930.

Received September 19, 1955.

THE RELATIONSHIP BETWEEN HUMOR AND THE EXPRESSION OF HOSTILITY¹

DONN BYRNE

Stanford University, Veterans Administration Hospital, Palo Alto

PSYCHOLOGICAL literature is marked with reports of sporadic attempts to harness humor as a psychodiagnostic tool. Before this goal can be attained, it is necessary to discover and understand the part, if any, that humor plays in personality dynamics. As a step in this direction, the present study was undertaken in order to relate both humor preferences and the understanding of humor to other behavioral characteristics.

The bulk of the experimental work dealing with humor has included the assumption, either explicitly or implicitly, that some sort of tension-releasing or drive-reducing process is involved in responding positively to humorous stimuli. This view of humor is consonant with psychoanalytic theory. Freud (6) postulated that the pleasure in what he calls tendency-wit results from the gratification of a tendency which would otherwise remain unfulfilled. Gratification is made possible because the censor is deceived temporarily by the humorous camouflage. Authors of later psychoanalytic writing similarly equate laughter and milder amusement with the release of repressed impulses and the consequent reduction of tension (2, 12, 14).

Proceeding from the idea that humorous material can serve to release repressed impulses or reduce drives, several attempts have been made to utilize humor as an indicator of these hidden forces. In an informal and exploratory study, Redlich, Levine, and Sohler (11) used cartoons as a device for obtaining clues to basic needs and conflicts among

neuropsychiatric patients and normals. Their observations led them to believe that a positive response to humor indicated the release of suppressed or repressed needs but that responses of indifference or dislike did not necessarily imply an absence of such needs. Murray (10), using paper-and-pencil questionnaires, assessed the aggressive sentiments and aggressive conduct of university students. He found that the former was highly related to the appreciation of disparaging jokes while the latter was not. From these results and the study of autobiographical material, he concluded that laughter at derisive jokes was the consequence of repressed hate.

A somewhat more inclusive approach is represented by Cattell, Luborsky, and Saunders (1, 3, 4, 9), who employed a factor-analytic technique. Their original intention was that of designing a humor test which would reveal an individual's major repressions by those jokes which he found amusing. To their surprise, a positive relation was consistently found between the vector measured by the joke cluster and the overt personality trait as defined by the Guilford-Martin questionnaires. Cattell resolved this apparent conflict between theory and findings by assuming that the deviation of overt and covert personality from the approved social norm is proportional. For example, the individual who frequently behaves in an overtly hostile manner also has more hostility hidden beneath the surface than the one who fails to manifest his hostility.

It appears that various investigators, approaching the study of humor with hypotheses based on Freudian theory, have found contradictory or inconclusive results, but have been loath to reject the theory.

The present study was designed to investigate the relationships among (a) behavior ratings of expression of hostility, (b) the extent to which hostile cartoons are judged to be amusing, and (c) the ability to recognize that the cartoons contain hostility. Freudian theory would seem to predict a negative correlation

¹From the Veterans Administration Hospital, Palo Alto, where this research was done as part of the work of the Research Unit in the Clinical Psychology Service. The author wishes to express his appreciation to the manager and staff of this hospital for making the present study possible, especially to the raters and judges: G. Brackbill, T. Coleman, M. Hansen, W. Hughes, W. Hunrichs, D. Johnson, L. Learned, R. St. Pierre, L. Ullmann, R. Worthington, and K. Wurtz. Special thanks go to S. Dean, R. McFarland, Q. McNemar, and P. McReynolds for their very helpful cooperation. Acknowledgment is also made to G. Berman and R. Miller, who were the other members of the Research Unit.

between expression of hostility and finding hostile cartoons amusing while previous research findings would tend to suggest a positive one. The purpose of this study was to help resolve this conflict.

METHOD

In order to secure a sizable sample of cartoons to serve as humorous stimuli, over 230 cartoons were removed from several back issues of the *Saturday Evening Post* and *Collier's*. These cartoons were presented to five psychologists and graduate psychology interns to be judged as expressing or not expressing interpersonal hostility. All five judges agreed on 80 of the cartoons, with 40 labeled hostile and 40 nonhostile. From this pool, 32 were selected in such a way as to form a hostile set of 16 matched as closely as possible in content with a nonhostile set of 16. The hostile and nonhostile cartoons were combined in random order to form the complete series.

The Ss consisted of male neuropsychiatric patients at the Veterans Administration Hospital, Palo Alto, California. On two separate admission wards, the psychiatrist, head nurse, and chief attendant independently rated a total of 89 patients on their typical manner of expressing hostility. All of the patients rated had been on the ward a minimum of six weeks and had been in close contact with the raters during this period. Each patient was placed in one of three groups according to definitions adapted from a rating scale developed by Schultz (12). These definitions were as follows:

"A patient classified as *overtly hostile* is one who 'acts out' aggressively in response to tension producing and frustrating situations. This aggression may be expressed either physically or verbally. The overtly hostile patient usually sees himself as being in the

position of having to defend himself by hostile and aggressive actions. He will often interpret members of the hospital staff as being very threatening and will react by physically or verbally attacking staff members directly or by running down the hospital policies, regulations, and activities in general.

"A patient classified as *covertly hostile* is one who expresses his hostility in a subtle, indirect, insidious, or passive manner. That is, regardless of whether the person is aware or unaware of the hostile implications of his behavior, he is in a position of not having to accept the responsibility for it. Some activities which might be seen as examples are: dumping cigarette ashes on the floor with an ash tray near by, frequently being late for ward activities and appointments, interrupting a person who is talking and then apologizing, or frequently being involved in accidents involving the belongings of others.

"A patient classified as *nonhostile* is one who does not fall in either the overtly hostile or covertly hostile categories. As far as can be observed, he does not act out his aggressive impulses in either overt or covert behavior. An extremely nonhostile patient might be described as the 'Caspar Milquetoast' type. Every effort is made to comply with all the demands and wishes of those around him, especially with the rules and regulations of the hospital."

An S was assigned to one of the three hostility groups if two of the three raters agreed on the grouping. Out of 89 patients rated, only six had to be eliminated because of failure to meet this criterion. From the remaining 83 patients, 45 Ss were randomly selected, with 15 in each of the three groups.

Table 1 indicates the characteristics of the 45 experimental Ss. Because the Ss of the nonhostile group differed markedly from those of the other two groups with respect to age, product-moment correlations between age and the two experimental measures were computed. Neither relationship was statistically

TABLE 1
AGE, LENGTH OF HOSPITALIZATION, AND DIAGNOSIS OF THE EXPERIMENTAL SUBJECTS

Characteristic	Experimental Group			
	Overtly hostile	Covertly hostile	Nonhostile	Total group
Age				
Range	24-72	22-55	31-72	22-72
Median	34	32	41	37
First and third quartiles	30 and 46	29 and 38	37 and 56	32 and 45
Length of Hospitalization (months)				
Range	3-79	4-54	2-386	2-386
Median	7	14	11	10
First and third quartiles	6 and 20	10 and 27	5 and 61	7 and 26
Diagnosis				
Schizophrenic disorder	5	11	4	20
Psychoneurotic disorder	2	1	2	5
Organic brain damage	2	1	1	4
Psychosis, unclassified	0	0	3	3
Character and behavior disorders	1	0	1	2
Paranoid disorder	1	0	0	1
Affective disorder	1	0	0	1
No diagnosis available	3	2	4	9

significant (Age with Hostile Humor Score, $r = -.05$, $p > .05$; Age with Hostility Judging Score, $r = -.25$, $p > .05$). In length of hospitalization there was a tendency for those who had been hospitalized the shortest time to express their hostility more overtly, but this variable was not significantly related to either experimental measure (Months of Hospitalization with Hostile Humor Score, $r = .04$, $p > .05$; Months of Hospitalization with Hostility Judging Score, $r = -.25$, $p > .05$). The best represented diagnostic category was that of schizophrenic reaction, with over half of these Ss subclassified as paranoid type. There was a preponderance of this category in the covertly hostile group, but the mean differences on the experimental scores between the schizophrenic Ss and those in the other diagnostic groups combined were not significant (Hostile Humor Scores, $t = .40$, $p > .05$; Hostility Judging Scores, $t = .51$, $p > .05$).

Each S, tested individually, was first asked to place the 32 cartoons in four equal piles according to how funny they were to him. The piles were labeled as "most funny," "next," "next," and "least funny." For the second task, each S was given the same cartoons in the original order and asked to divide them into two equal piles. In one pile were placed those cartoons in which "insults, anger, teasing, or meanness" was expressed, and in the other pile those in which these factors were not present. In effect, he was informed that half the cartoons contained hostile material; his task was to divide the 32 cartoons appropriately. The entire testing procedure lasted from 20 to 40 minutes. The reliability of the hostility-judging task performed by the Ss was .86 as computed by the split-half method corrected by the Spearman-Brown formula.

RESULTS

The three groups were compared on their ratings of the funniness of the hostile cartoons. For every S, three points were assigned for each hostile cartoon rated most funny, two points for each rated next most funny, etc. The total number of these points for each S constitutes his Hostile Humor Score. The means and standard deviations of the resulting distributions are given in Table 2. A simple one-way analysis of variance of the group means failed to show significant differences ($F = 2.27$, $p > .05$). However, when the Ss were divided into two groups, those who express hostility either overtly or covertly and those who do not express hostility, a significant difference between groups was found ($t = 2.11$, $p < .05$). Those Ss who express hostility find hostile cartoons funnier than do those Ss who do not.

Another complication arose in analyzing the second task. It was found that the number of cartoons correctly judged to express or not to express hostility (designated as the Hostility

Judging Score) correlated significantly with IQ as estimated by four subscales of the WAIS ($r = .75$, $p < .01$). These intelligence scores were available for only 10 of the Ss in each group. It should be noted that the correlation between the Hostile Humor Score and IQ was not significant ($r = -.10$, $p > .05$).

Table 3 reports the means and standard deviations of the Hostility Judging Scores of the three groups. An analysis of variance revealed no significant group differences in the mean number of cartoons judged correctly ($F = 1.28$, $p > .05$). With intelligence controlled by the covariance method, as shown in Table 4, the three groups still did not differ

TABLE 2
COMPARISON OF GROUP DIFFERENCES IN THE HUMOR RATINGS OF HOSTILE CARTOONS

Group	Humor Ratings of the Hostile Cartoons		
	N	M	SD
Overtly hostile	15	25.33	2.33
Covertly hostile	15	24.93	3.57
Nonhostile	15	23.00	3.27
Overt and covert groups combined	30	25.13	3.02

TABLE 3
COMPARISON OF GROUP DIFFERENCES IN JUDGING THE HOSTILE CONTENT OF CARTOONS

Group	Correct Judgments of Cartoon Content		
	N	M	SD
Overtly hostile	15	22.93	5.31
Covertly hostile	15	21.47	4.92
Nonhostile	15	19.47	6.83

TABLE 4
COMPARISON OF GROUP DIFFERENCES IN JUDGING THE HOSTILE CONTENT OF CARTOONS WITH INTELLIGENCE CONTROLLED BY COVARIANCE

Group	Correct Judgments of Cartoon Content			Estimated IQ	
	N	M	SD	M	SD
Overtly hostile	10	23.80	5.83	104.30	16.08
Covertly hostile	10	21.20	4.75	92.40	10.47
Nonhostile	10	18.20	5.69	95.50	16.33
Overt and covert groups combined	20	22.50	5.47	98.35	14.82

significantly on the task ($F = 2.91, p > .05$). Again, however, when the *Ss* are divided into expressers and nonexpressers of hostility, significant differences occur ($F = 5.82, p < .05$) when IQ is controlled by covariance. Those *Ss* who express hostility are better able to recognize hostility in cartoons than are the *Ss* who do not express it.

The only other finding of note was that the Hostile Humor Scores and the Hostility Judging Scores were not significantly related ($r = .09, p > .05$) unless the effects of IQ were partialled out ($r_{12 \cdot IQ} = .37, p < .05$). In view of the previous findings, this result is not surprising.

DISCUSSION

The findings reported here appear to corroborate previous experimental results and to contradict psychoanalytic theory in that a positive relationship was obtained between the type of humor found amusing and overt personality characteristics. Cattell's suggested resolution would explain the present findings by hypothesizing that those *Ss* who express more hostility also had more underlying, unexpressed hostility, and it was the latter factor which determined the response to the cartoons. Other possibilities exist, however, and it seems reasonable to explain the findings without depending on postulated repressed needs.

Another thesis to account for the rather consistent experimental results can be stated in Hullian terms. A positive response to a particular humorous theme may be conceived as a function of both drive and habit strength and not, as other investigators have hypothesized, a function of the strength of repressed drives only. From this viewpoint, the appreciation of hostile cartoons is seen simply as an expression of hostility by individuals who also express hostility in other situations. From Hull's views (7), it follows that individuals with strong aggressive drives who experience drive reduction by expressing hostility in a given situation have increased habit strength for the expression of hostility in that situation. Since Feshbach (5) found that aggressive drive is partially reduced by its expression in fantasy material, it seems likely that the potential drive-reducing situations include fantasy and humor. An additional postulate (7) is that habits are evoked, on the basis of similarity,

by stimuli other than the original ones. Therefore, the situations in which hostility is expressed could be arranged on a gradient of generalization. As an example to illustrate both postulates, it would be predicted that individuals who were allowed to express hostility toward their fathers without punishment would express hostility toward other authority figures, would enjoy stories in which such hostility was expressed, and would be amused by humorous material with this same theme.

Another question, and one not answered by the present study, is whether *Ss* who express hostility necessarily possess the highest aggressive drives. One possible answer comes from Holzberg, Bursten, and Santiccioli (8), who found that *Ss* who express either more or less than average aggression in fantasy have strong aggressive tension. Perhaps the non-hostile *Ss* in the present study had as strong aggressive drives as the overtly and covertly hostile ones. Systematic behavior theory would account for the difference in present behavior on the basis of punishment following the early aggressive behavior of the non-hostiles.

While the instrument reported in this study yielded significant group differences, its value as a predictor of individual behavior is low. However, these various findings and tentative generalizations suggest that it may be possible to construct longer and greatly refined humor tests in which the degree of positive response to hostile humor would measure habit strength for the expression of hostility. A second suggestion is that the degree of either positive or negative response to the humor would be a measure of drive strength. Further, one might logically expect these same relationships to hold for other drives besides aggressive or hostile ones, so other appropriate humor tests could be devised. One last possibility is that humor tests could be made very specific as indicators of particular habits in particular situations.

The finding that those *Ss* who express hostility are better able to recognize it in the cartoons remains to be explained. If both hostile and nonhostile *Ss* have strong aggressive drives, it could be postulated that the former are more sensitive to stimuli which offer a reduction in their drive state, whereas

the latter are less motivated to recognize aggressive stimuli because response to such stimuli has not been rewarded in the past. If the nonhostile Ss are actually avoiding aggressive cues (instead of merely being less sensitive), this phenomenon amounts to what other investigators have designated as perceptual defense against threatening stimuli.

The positive correlation between intelligence and the ability to identify correctly the hostile and nonhostile cartoons points to the possibility of utilizing humor to a greater extent than is presently done in the measurement of intelligence. The task involved in this research required the understanding of either the point of each cartoon or the interpersonal situation depicted. In other words, the Hostility Judging Scores seem to represent, in part, an index of the ability to carry on abstract thinking. The use of cartoons as an intelligence test subscale may prove useful if they provide a reliable and economical measure. It should be noted, though, that with the cartoons used in this study, a ceiling IQ of about 120 was found. Consequently, humor as an intelligence measure may only be useful in rough screening devices. Since both intelligence and personality dynamics were found to be related to the task of identifying hostile cartoons, any successful attempt to construct either a diagnostic device or an intelligence scale with humor as a base must avoid the confounding of these two factors.

A word of caution about the experimental results seems in order. It is undoubtedly desirable to present any experimental findings simultaneously with cross-validation material. At the very least, there should be some assurance from the investigator that the measures which he is using show some stability over time. In practical situations the fulfillment of these ideals is not always immediately possible; the present study is an example. Nevertheless, the experimental results were statistically significant, tend to be compatible with previous findings, and may serve to stimulate further work in the area.

SUMMARY AND CONCLUSIONS

This study was undertaken in order to help resolve the conflict between Freudian theory and previous experimental findings by ex-

ploring the relationships among (a) the expression of hostility in behavior, (b) the extent to which hostile cartoons are judged funny, and (c) the ability to recognize hostility in cartoons.

Three groups of male neuropsychiatric patients were rated as expressing hostility overtly, covertly, or not at all. They were asked to judge the funniness of 32 cartoons and also to indicate which of the cartoons expressed hostility and which did not. It should be emphasized that the generality of these findings can not be known until the study is cross validated with other Ss.

It was found that:

1. Those Ss who frequently express hostility, either overtly or covertly, find hostile cartoons significantly more amusing than do those Ss who fail to express hostility.
2. There is a significant positive correlation between ability to differentiate hostile and nonhostile cartoons and estimated IQ, but there is no significant correlation between IQ and finding the hostile cartoons amusing.
3. Those Ss who frequently express hostility, either overtly or covertly, are significantly better able to differentiate hostile and nonhostile cartoons than are those Ss who fail to express hostility, if the effects of intelligence are statistically controlled.
4. There is a significant positive correlation between ability to differentiate hostile and nonhostile cartoons and finding hostile cartoons amusing only if the effects of intelligence are partialled out.

REFERENCES

1. ANDERSON, H. H., & ANDERSON, GLADYS L. *An introduction to projective techniques*. New York: Prentice-Hall, 1951.
2. BRILL, A. A. The mechanism of wit and humor in normal and psychopathic states. *Psychiat. Quart.* 1940, **14**, 731-749.
3. CATTELL, R. B., & LUBORSKY, L. B. Personality factors in response to humor. *J. abnorm. soc. Psychol.*, 1947, **42**, 402-421.
4. CATTELL, R. B., & SAUNDERS, D. R. Inter-relation and matching of personality factors from behavior rating, questionnaire, and objective test data. *J. soc. Psychol.*, 1950, **31**, 243-260.
5. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3-11.
6. FREUD, S. *The basic writings of Sigmund Freud*. New York: Random House, 1938.
7. HILGARD, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.

8. HOLZBERG, J. D., BURSTEN, B., & SANTICCIOLI, A. The reporting of aggression as an indication of aggressive tension. *J. abnorm. soc. Psychol.*, 1955, 50, 12-18.
9. LUBORSKY, L. B., & CATTELL, R. B. The validation of personality factors in humor. *J. Pers.*, 1947, 15, 283-291.
10. MURRAY, H. A., JR. The psychology of humor. *J. abnorm. soc. Psychol.*, 1934, 29, 66-81.
11. REDLICH, F. C., LEVINE, J., & SOHLER, T. P. A mirth response test: Preliminary report on a psychodiagnostic technique utilizing dynamics of humor. *Amer. J. Orthopsychiat.*, 1951, 21, 717-734.
12. REIK, T. *Listening with the third ear*. New York: Farrar, Straus, 1948.
13. SCHULTZ, S. D. A differentiation of several forms of hostility by scales empirically constructed from significant items on the Minnesota Multiphasic Personality Inventory. Unpublished doctor's dissertation, The Pennsylvania State Univer., 1954.
14. SKELLS, D. The function of humor in three Nez Perce Indian myths. *Amer. Imago*, 1954, 11, 249-261.

Received September 19, 1955.

DISTRIBUTIONS OF TRAITS IN CURRENT Q-SORT METHODOLOGY¹

AUSTIN JONES²

Central State Hospital, Indianapolis

A NUMBER of investigations within the last few years have used the method referred to as Q sort for research in the areas of personality and psychotherapy. The method has been drawn from the group of related inverted factor-analytic procedures known as Q technique, dealt with most extensively by Stephenson (11). In the initial study by Hartley (5), the Q sort was used to study changes in the client's "self concept" and "ideal self" and the relationship between them in the course of therapy. Butler and Haigh (1), in a recent study, investigated the same relationship. Rudikoff (10) studied the relationships between concepts of self, "ordinary person," and the "ideal" at several points in time relative to the therapeutic process. In addition to these researches, there have been a number of others using Q sort in the investigation of closely related topics (3, 4, 6, 8, 9, 12).

Common to all of the studies cited has been the characteristic Q-sort procedure which requires the *S* to sort a number of statements of self-reference into a series of piles along a continuum of appropriateness or accuracy of self-description, from those that are "least like" him to those that are "most like" him. The number of items sorted into each pile is specified in such a way that the frequency distribution of items in the piles along the continuum approximates the normal distribution. The items may be selected in a number of ways but in each case are presumed to sample randomly a previously defined "trait universe."

The practice of forcing a distribution (in this case, a quasi-normal distribution) was adopted for several reasons. Perhaps the more important ones are that such a procedure acts to reduce response sets and that variance in the

responses is automatically obtained. It is also a procedure of some statistical convenience, as the means and standard deviations of all *Ss'* distributions become automatically the same and thus greatly simplify computation of the product-moment correlations involved in each case. Some limitations inherent in this procedure have already been noted. Cronbach, for instance, has pointed out the loss of information that may occur when individual differences in shape of distribution are suppressed. Forced-distribution procedures, he comments, "... may be psychologically indefensible, if there is reason to think that persons differ in their variability over traits" (2, p. 379).

Since Cronbach's note of caution there has been little if any change in the Q-sort methodologies in published research. The studies cited earlier, with the exception of Hartley's, have all appeared since Cronbach's article, thus leading to the impression that the authors have taken the position that there are no significant differences in the shape of distributions of traits among individuals.

The characteristic practice of forcing a quasi-normal distribution would seem to imply further that the characteristic shape of distribution of traits within individuals is also a quasi-normal distribution. These statements do not apply only to the set of trait statements but also, since they are presumed to be random samples from the trait universe, to the distribution of the trait universe as well.

The formulation that has been inferred from current Q-sort procedures may be stated more specifically as follows: The distribution of the degree of development of the universe of traits within individuals is a quasi-normal distribution; although relative development of different traits may vary from time to time, the degree of development of traits within individuals remains always a quasi-normal distribution. In other words, the strength or predominance of a given trait may wax and wane in the individual's life, but the distribution of strength of traits (operationally, the degree of their self-descriptiveness) remains constantly quasi-normal.

¹ The author wishes to express his appreciation to Dr. Morton Wiener, Chief Psychologist of Central State Hospital, for his advice and encouragement in the execution of this study and to Dr. C. L. Williams, Superintendent of Central State Hospital, and Dr. Ruth B. Hines, Chief Psychologist of Veterans Administration Hospital, Indianapolis, for permission to utilize patients as research subjects. The author is indebted to James Norton of the Statistical Laboratory, Purdue University, for statistical help and advice.

² Now at The University of Rochester.

The purpose of the present study is to subject this implicit theoretical formulation to experimental verification and to assess the appropriateness of the corresponding statistical procedure of the forced quasi-normal distribution. (It should be noted that there are at least two antecedent assumptions of Q-sort procedures, each of which may merit study. One is the assumption that the Q-sort items randomly sample the trait universe. The second is the assumption that the Ss are able to report meaningfully or accurately their perception of the degrees of self-reference of the Q-sort items. For the purposes of this study, these assumptions are provisionally accepted.)

Within the framework of the theoretical position stated above, it may be hypothesized that individuals, given statements which are presumed to sample the total population of traits and allowed to indicate what they feel to be the relative appropriateness of self-description of the statements, will assign values in such a way that their frequencies will closely approximate the normal distribution. It is with the testing of this hypothesis that the present paper is concerned.

A "free-sort" procedure, i.e., one in which the distribution is not specified, would be expected to give evidence of, among other things, the range of meanings with which the individual is able to interpret himself in relation to the world or by which he perceives himself. This is inherent in the task of rating the appropriateness of self-description of items. There seems to be no readily apparent reason why individual differences would not manifest themselves along this as well as other personality variables. Thus, the first of the two predictions of this study is: Ss will show significant variation among themselves in regard to shape of distribution sorted, and, concomitantly, Ss will show significant variation in shape of curve from the quasi-normal curve predicted by the theoretical position described earlier.

Maladjusted individuals have been described from time to time as suffering from distortions and disturbances in the variety of meanings with which they can interpret themselves and situations about them. A recent statement of this view has been made by Osgood and Luria (7). It seems consistent with such theory to hypothesize that maladjusted individuals in a

"free-sort" procedure would manifest a tendency to deny appropriateness of self-description of items and to dichotomize items in such a way that only a few degrees of appropriateness of self-description are admitted. The second prediction of this study, then, may be stated formally as: Maladjusted Ss will show greater deviation in shape of distribution from the normal distribution than will less maladjusted or "normal" Ss, with greater degrees of maladjustment showing greater degrees of deviation.

METHOD

Subjects. Three groups of ten Ss each were chosen to represent three rather widely varying degrees of personality integration or adjustment. The first group was drawn on a volunteer basis from the male students taking a required introductory college course in psychology. The second group was drawn randomly from the population of male neurotic patients receiving treatment at a Veterans Administration hospital. Patients with organic involvement or mental deficiency, with prior history of psychosis, over 36 years of age, or receiving insulin or electroshock within the six weeks prior to testing were excluded. Anxiety neurosis was the primary classification. The third group was drawn from the population of male patients at a state mental hospital, each of whom would be commonly regarded as psychotic. These patients were selected as they were referred routinely for psychological evaluation. Patients grossly out of contact, over 36 years of age, with organic involvement or mental deficiency, and those having received insulin or electroshock treatment within the previous six weeks were excluded as Ss. Schizophrenic disorders were predominant.

The groups were equated for age and education as closely as the populations permitted. The mean ages for the college, neurotic, and psychotic groups were 24.0 years, 28.8 years, and 29.7 years respectively. The mean educational level for the three groups (highest grade completed) was 12.8, 11.3, and 12.2.

Materials. The materials used were the 100 items of the SIO Q sort developed by Butler and Haigh (1). This particular set was selected because of its frequent and systematic use in recent research. The items are statements of self-reference typed on individual cards. (Sample items: "I am a dominant person." "I am liked by most people who know me." "It's hard to control my aggression.")

Procedure. All Ss participated on a volunteer basis, their right to decline having been clearly expressed. The Ss were informed that the results of their performance would have no bearing on course work in the case of the college Ss, or upon hospitalization or type of treatment in the case of the neurotic and psychotic groups. In the course of testing, one S from the neurotic population and two from the psychotic population declined to complete the task and were necessarily excluded as Ss.

All Ss were seen individually by the author. The Ss were asked to sort the 100 items as to their degree

of appropriateness of self-description on a 9-point scale of equal-appearing intervals. Blue cards numbered one through nine were placed in a row on the table surface to indicate the ascending order of points on the continuum. The instructions were those used by Butler and Haigh and others with the SIO Q sort, plus additional instructions for the equal-appearing intervals aspect of the task. A standardized set of gestures was used concurrently with the verbal instructions for the purpose of making the scaling procedure as clear as possible. The *E* presented the following instructions orally. The portion italicized is the basic instruction used by Butler and Haigh.

"Here is a set of cards which have on them statements about how people think and feel. Your task is to *sort these cards to describe yourself as you see yourself today from those that are least like you to those that are most like you*. Sort them into nine piles. In the first pile put the statement—or statements—that describes you least well, that is least like you. In the ninth pile put the statement—or statements—that describes you best, that is most like you. In the piles in between, arrange the other statements so that each pile describes you better and better, and so that the distances between them are even. That is very important. Try to make the distances between the piles as even as possible, so that the statements in this pile [the second] are more like you than in this one [the first]—and those in the next pile [the third] are by the same amount more like you than in this one [the second]—and those in this pile [the fourth] are by the same amount more like you than in this one [the third]—and so on. So that the increase from pile to pile is the same. Do you understand?"

"Now in each pile you must put at least one card. Beyond that you may put as many or as few as you wish in any pile—even none."

"Are there any questions?"

Where information regarding reading ability was incomplete, *Ss* were requested to read sample items aloud. Two *Ss* from the psychotic population were found to have only minimal comprehension of the items and were thus excluded from the study. The *Ss* were informed that they might ask questions as they went along. By far the majority of the questions concerned the meaning of individual words. In such cases *E* encouraged verbalization of *S's* impression of the word. When no reaction was forthcoming or when the word was badly misunderstood, *E* supplied as standardized a meaning as possible.

In order to ascertain whether or not the *Ss* were, in fact, using a 9-point scale of equal-appearing intervals, *E* conducted a brief postexperimental inquiry with each *S* after he had completed the sorting. The following question or a variation of it was asked: "What was it about the task that you found most difficult?" In only one case was the equal-appearing intervals aspect cited. With further inquiry, this *S* indicated that while he found it the most difficult part, he nevertheless felt that he had mastered it adequately. The other *Ss* were then queried specifically concerning the equal-appearing intervals aspect. All *Ss* reported their confidence on this point. When *E* still felt any doubt, *S* was further encouraged to verbalize the scaling principle as a final check. In each case he responded adequately.

RESULTS

A comparison of the shape of each *S's* distribution with that of the approximate normal distribution customarily imposed in use of the SIO Q sort was made by use of the chi-square test of goodness of fit. The number of items falling at each of the nine classes for each *S's* sorting constituted the obtained frequencies. The numbers of items customarily required in each class for the approximate normal distribution were the "expected" frequencies. Significant differences between obtained distributions and the approximate normal distribution were found for all *Ss* in each of the three groups; *p* in all cases was less than .01.

The composite (or average) distributions for each of the three groups are shown with the approximate normal distribution in Fig. 1. Differences among the three curves were studied by the analysis of variance which is summarized in Table 1. The proportion of the 100 Q-sort items sorted into each of the classes along the continuum of appropriateness of self-description constituted the scores for each *S*. The arc-sine transformation was applied to the scores. Bartlett's test was computed on the transformed scores and the homogeneity-of-variance assumption supported. Analysis over the three curves failed to show significant differences. Separate *F* tests, however, comparing each curve with each of the others showed significant dissimilarity between the

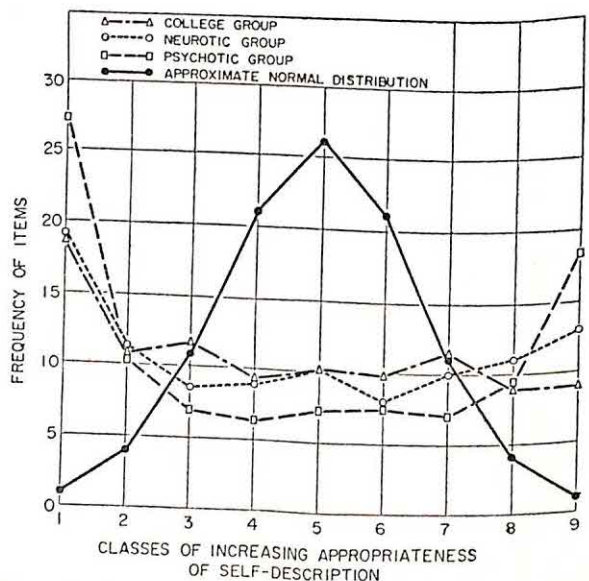


FIG. 1. COMPOSITE DISTRIBUTIONS OF THE THREE EXPERIMENTAL GROUPS SHOWN WITH THE APPROXIMATE NORMAL DISTRIBUTION

TABLE 1
ANALYSIS OF VARIANCE OF PROPORTION OF ITEMS
PLACED AT EACH CLASS INTERVAL
(Data in the form of arc-sine transformation
for proportions)

Source	df	MS	F
Group \times class	16	7790.4	1.33
Pooled Ss \times class	216	5841.7	
Group (excluding neurotic) \times class	8	13596.7	2.32*
Pooled Ss (excluding neurotic) \times class	144	5851.6	
Group (excluding college) \times class	8	7243.0	1.23
Pooled Ss (excluding college) \times class	144	5892.2	
Group (excluding psychotic) \times class	8	2530.5	
Pooled Ss (excluding psychotic) \times class	144	5788.2	

* Significant at the .05 level.

curves for the college and for the psychotic groups ($p < .05$). The objection might be raised that declaring such an outcome "significant" is, to an extent, "capitalizing on chance" because of the interdependence of the three F tests. In this case, however, the rank relationship obtained among the three curves is a specifically predicted one—that maladjusted Ss will show greater deviation in the shape of their distribution from the approximate normal distribution than will less maladjusted Ss, with greater degrees of maladjustment (neurotic, psychotic) showing greater degrees of deviation. Thus, it would seem that the bias inherent in ascribing significance to this comparison is not as great as might otherwise be the case.

The comparison of curves for Ss within groups was made by contingency chi squares, as indicated in Table 2. For each of the three

TABLE 2
CONTINGENCY CHI SQUARES COMPARING SUBJECTS'
CURVES WITHIN EACH EXPERIMENTAL GROUP

Group	df	χ^2
College	72	1986880.0*
Neurotic	72	30751.8*
Psychotic	72	356118.0*

* Significant at .001 level.

groups the chi square was based on a work table of ten Ss' sorting of items into nine class intervals. Significant differences were found among the item distributions of Ss comprising each of the experimental groups, indicating that none of the groups could be characterized by a single distribution shape with only minor subject-to-subject variation from it. Further analysis, however, may yield two or three composite curves for a given group which together subsume a very high proportion of the total variance. In general, two major shapes are apparent, those skewed to the right and those skewed to the left.

DISCUSSION

The hypothesis drawn from current Q-sort procedures that there is a quasi-normal distribution of the degree of development of traits within individuals is not supported by the findings of this study, as was predicted. The application of chi square to a comparison of obtained curves with the approximate normal distribution resulted, as noted earlier, in significant differences for all Ss, regardless of group membership. The distribution of the universe of traits within individuals, assuming that the SIO Q-sort items are a random sampling of that universe, is clearly indicated to be non-normal and as varying from individual to individual. There is no evidence for a single characteristic distribution of any kind.

It may be worth noting that the lack of congruence between the forced quasi-normal distribution and the "free-sort" distributions—which may be regarded as more accurate estimates of the distribution of the trait universe—indicates distortions in the absolute values of product-moment correlations computed between Q sorts. Thus far, this possibility would not seem a major problem because Q-sort researches have been typically concerned with the identification of *direction of change* in certain correlations, an aspect which is less affected by such a lack of congruence.

The second prediction, that deviation from normality of distribution increases with increasing maladjustment, received partial verification by the results of this study. The three composite curves—college, neurotic, and psychotic—seem increasingly U-shaped in that order (Fig. 1). Although this order corresponds to the one predicted, only the difference between the college and psychotic groups is

statistically significant. These findings suggest that the most maladjusted (specifically, psychotic) Ss perceive a more restricted, limited range of intensity of references to the self. Such individuals in general react to self-reference material as either very "true" or very "untrue" of them. Responses indicating an acceptance of variable feelings or behavior, such as statements sorted near the middle of the scale, are much less common. To put it in information theory terms, it appears that more maladjusted persons attempt to organize the world in concepts of low information and high redundancy or predictability. The maladjusted individual appears less able than others to tolerate concepts of high variability or low predictability, at least in relation to himself. In contrast, the college Ss, while still deviating significantly from the normal distribution, sorted many more statements into the center piles of the scale, thus accepting more readily concepts of self as more variable and of lower predictability. As indicated, the neurotics responded in the direction of the psychotics but did not differ significantly from the college group.

The results suggest distinct advantages inherent in the "free-sort" procedure employed in this study. In general, the use of this procedure seems to preserve the great amount of information that is now lost in forced distribution methods. The possibility of differences in shape of distribution between groups of varying degrees of maladjustment suggests that meaningful personality correlates of shape of distribution might be determined. As noted earlier, two characteristic shapes of curve seem to emerge on inspection, one skewed to the right and one skewed to the left. These distributions may be thought of as representing, respectively, primarily repressive and primarily sensitized orientations. In the one case, Ss have tended systematically to deny the self-reference of the statements; in the other, they have tended to perceive the statements predominantly highly applicable to themselves. Once the stability of the shapes of "free-sort" distributions is known, the significance of temporal changes in shape of distribution under various personality or therapeutic conditions might be profitably studied. It would seem, then, that "free-sort" procedures, through their property of retaining information, would lead to an extension and continued develop-

ment of Q technique and related correlation-between-persons methods.

SUMMARY

Current Q-sort methodology, through its characteristic use of forced, quasi-normal distributions, appears implicitly to support the theoretical position that the characteristic shape of distribution of traits within individuals is a quasi-normal one and that there are not significant differences in shape of distribution of traits among individuals. It was the purpose of the present study to assess experimentally this theoretical formulation. Three groups of Ss, college, neurotic, and psychotic, were administered the SIO Q sort with instructions to sort the items on an equal-appearing-intervals scale with the same range as that customarily used but with an unspecified or "free" distribution rather than a forced, quasi-normal distribution.

The results showed that each S's "free-sort" distribution varied significantly from the approximate normal distribution customarily imposed in use of the SIO Q sort and that within each of the groups, Ss' shapes of distributions varied significantly from each other. Composite curves were constructed for each of the three groups. An analysis of variance over the three curves failed to show significant differences. Separate *F* tests comparing each curve with each of the others showed significant dissimilarity only between the curves for the college and for the psychotic groups, with the psychotic group showing the greater deviation from the approximate normal distribution. The interdependence of the three *F* tests, however, raises some question as to the actual significance of the finding.

It was concluded that current forced-distribution procedures result in a significant loss of information which may be retained by use of "free-sort" procedures of the type described here. The possibility of differences in shape of distribution between groups of varying degrees of maladjustment suggests that meaningful personality correlates of distribution shape might profitably be investigated.

REFERENCES

1. BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts. In C. R. Rogers & Rosalind Dymond

- (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
2. CRONBACH, L. J. Correlation between persons as a research tool. In O. H. Mowrer (Ed.), *Psychotherapy: theory and research*. New York: Ronald, 1953.
3. GRUMMON, D. L. Personality changes as a function of time in persons motivated for therapy. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
4. HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal self in relation to personality adjustment. *J. consult. Psychol.*, 1954, **18**, 215-218.
5. HARTLEY, MARGARET, A Q technique study of changes in the self-concept during psychotherapy. Unpublished doctor's dissertation, Univer. of Chicago, 1951.
6. NUNNALLY, J. C. An investigation of some propositions of self-conception: the case of Miss Sun. *J. abnorm. soc. Psychol.*, 1955, **50**, 87-92.
7. OSGOOD, C. E., & LURIA, Z. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49**, 579-591.
8. ROGERS, C. R. The case of Mrs. Oak: a research analysis. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
9. ROGERS, C. R. The case of Mr. Bebb: the analysis of a failure case. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
10. RUDIKOFF, E. C. A comparative study of the changes in the concepts of the self, the ordinary person, and the ideal in eight cases. In C. R. Rogers & Rosalind Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. of Chicago Press, 1954.
11. STEPHENSON, W. *The study of behavior: Q-technique and its methodology*. Chicago: Univer. of Chicago Press, 1953.
12. TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 205-209.

Received September 28, 1955.

MOTIVATION AS A VARIABLE IN WORK-PARTNER SELECTION¹

ELIZABETH G. FRENCH

Personnel Research Laboratory, Air Force Personnel and Training Research Center

THE hypothesis tested in the present experiment is that the behavior of an individual making a work-partner choice between a competent nonfriend and a less competent friend can be predicted from knowledge of his goal orientation. Specifically it was hypothesized that Ss with high achievement motivation would select the competent nonfriend; those with high affiliation motivation, the less competent friend. The Ss high in both motivations were expected to show evidence of conflict.

METHOD

In a group session, friendship ratings and achievement- and affiliation-motivation scores were obtained from the Ss and the importance of doing well on a test of "concept formation" was stressed. Then the Ss were divided into groups of four on the basis of the friendship ratings. Each group was made up of three mutual friends and an S who was not considered a friend by the other three. The "concept formation test" was explained to each group in turn and the four Ss worked on it individually. The nonfriend (labeled SS) was permitted to succeed; the other three were made to fail. On the expiration of the time limit, the Ss were told that they were to work on a similar test in pairs and asked to write down their choice of a work partner. The choices of the three unsuccessful Ss comprised the experimental data.

Subjects. The Ss were airmen in their seventh week of basic training at Lackland Air Force Base. A total of 204 men, members of seven flights, took part. The data are based on 137, as explained above.

Friendship ratings. Twenty-eight or 32 members of each flight reported as a group. They were told that they were to take part in an investigation of factors related to formation of friendships in the Air Force. Each man was given a roster containing the names of all the men present and asked to rate them 1, 2, or 3 according to whether he liked the man in question and considered him a friend, 1, didn't know him very well or was indifferent to him, 2, or disliked him, 3. In addition they checked their three best friends and any men they particularly disliked. The flight is the training group into which the men are organized as soon as they arrive on the base. They live, train, study, and work

together in virtual isolation from other groups during the training period. It is reasonable to assume that by the end of seven weeks patterns of friendship would be stable and meaningful.

Motivation measurement. After the Ss had completed the friendship ratings, they were told that the second purpose of the investigation was to determine the relative ability level of men entering the Air Force at different times. It was explained to them that it was very important for them to do as well as possible on the "concept formation" test they were soon to take. This was done to provide an achievement orientation similar to the affiliation orientation presumably resulting from the making of the friendship ratings. Then the motivation measurement device was presented as a task unrelated to either of the main purposes of the experiment. This instrument has been described in detail elsewhere (1, 2)². Called a "Test of Insight" it is presented to the S as a measure of his ability to understand the behavior of others. The test items are single-sentence descriptions of behavior supposedly characteristic of an individual with whom S can identify. (Example: "Tom always lets the 'other fellow' win.") The S is to "explain" the behavior. The number of achievement-relevant and affiliation-relevant responses make up the scores.

Composition of work groups. While the Ss were taking the Test of Insight, the friendship ratings were examined and each S was assigned to a work group with the following considerations in mind. Preliminary investigation had revealed that if best friends—Ss who rated each other 1 with a check—were put in the same group they invariably chose each other. Also if the nonfriend were either strongly disliked by the other work-group members, or very unpopular in the flight as a whole, he was never chosen. The artificially created necessity to succeed on the task could not be made as strong as the more extreme likes and dislikes. Therefore best friends were not grouped together nor was the SS for any group one whom any other member had checked as especially disliked. Secondly, approximately half the Ss failed to use the 3 category on the friendship scale at all. In these cases their 2 ratings were considered to indicate lack of friendship and the SS for their groups were drawn from those Ss so rated. Finally, the SS was selected to have an Armed Forces Qualifications Test score equal to or higher than the other members of the group in order to give plausibility to the success and failure manipulations. Each group, then, was composed of three Ss who had rated each other 1 without checking the rating, and one S with a high AFQT score who was rated 3 (or 2) by the other three Ss. In 16 of the 51 groups it was possible to satisfy the criteria for only two of the three critical Ss. The data for the third Ss in these groups, of course, were not used.

¹ This investigation was carried out under the Air Force Personnel and Training Research Center in support of Project No. 7704. Permission is granted for reproduction, translation, publication, and use or disposal in whole or in part by or for the United States Government. Portions of this paper were read at the 1955 meeting of the American Psychological Association.

² Copies of a paper describing the test and its development and directions for scoring may be obtained from the author.

The experimental situation. The experimental room was equipped with four individual booths so arranged that the Ss could not see each other but all could see *E* and a blackboard. The four Ss were seated in the booths and *E* wrote each S's name on the board. The *E* then reminded the Ss of the importance of succeeding on the task and gave them their instructions. The task involved sorting 30 cards containing various combinations of straight and curved lines, segments of geometric figures, and complete figures into categories. There was no logically compelling solution and almost any sort could be called "correct." The Ss, however, were told that there was only one correct answer. The instructions were:

"These cards contain designs which can be sorted into a definite and specific number of categories. Your task is to decide what the categories are and arrange the cards accordingly. There is only one right answer but you may have more than one trial. When you think you have the solution raise your hand and I will check it. You can offer as many solutions as you like until the time is up. The time limit will be five minutes after the first man gets it right."

The first solution offered by the SS was accepted as correct, his success was announced, and his name was circled on the blackboard. All solutions offered by the other Ss were called incorrect. Five minutes after the SS had finished, time was called. The *E* then explained that the Ss were to work on a similar task in pairs and asked each man to write down the name of the man with whom he would like to work. In order to permit an expression of conflict, the Ss were told that if there were two men with whom they would like to work equally well they might write down both names.

The Ss were paired and worked on a second set of cards but the critical part of the experiment ended with the choice of a partner.

RESULTS AND DISCUSSION

The achievement-motivation scores ranged from 0 to 15 with a mean of 5.65 and an *SD* of 2.98. The median fell between 5 and 6. For affiliation motivation the range was 0 to 15 with a mean of 4.55, an *SD* of 2.48, and the median between 4 and 5. The Ss were divided into four motivation groups on the basis of these scores: those with achievement-motivation scores above the median and affiliation below (High Achievement); those with affiliation above the median and achievement below (High Affiliation); those with both means above the median (Both High); and those with both means below (Both Low). Since the two sets of scores are relatively independent ($r = .17$) the number falling into each group (Table 1) is approximately equal.

The choices were tabulated in five categories: Choice of the SS only; choice of the SS first, then a friend (SF); choice of a friend

TABLE 1
FREQUENCY OF CHOICES IN EACH CATEGORY BY
MOTIVATION GROUP

	High Achievement	Both High	High Affiliation	Both Low	Total
Success-subject only	12	4	0	2	18
Success-subject—friend	7	4	0	1	12
Friend—success-subject	5	14	4	5	28
Two friends	2	3	11	3	19
One friend	9	8	18	25	60
Total	35	33	33	36	

and then the SS (FS); choice of two friends (FF); and choice of a single friend (F). The two orders of choice of SS and friend were separated on the assumption that the first-listed man might be the preferred one even though the instructions were to list two only if they were equally preferred. To the extent that this might be true the SF order was expected to be related to relatively higher achievement and lower affiliation motivation than the FS order.

Inspection of Table 1 reveals large differences in the frequency with which the various kinds of choices were made. These frequencies are significantly different from chance expectancies and well beyond the .001 level (chi square is 55.68). The large number of F choices suggests that in spite of the attempt to decrease the importance of the friendship variable in setting up the groups, friendship remained more important than task success.

Even with the widely divergent frequencies in the choice categories, the choices of the four motivation groups conformed to prediction (Table 1). One third of the High Achievement group chose the SS and over half made SS or SF choices. This group made two thirds of the total SS choices and over half of the SF choices. More than half of the Both High group made either SF or FS choices. These Ss accounted for half the FS choices. The High Affiliation group made no SS or SF choices and only four FS. All their other choices were either FF or F. This group made over half the FF choices and appears to be the only one showing much evidence of conflict over choosing between two friends. The Both Low Ss made a few choices in the first four categories,

TABLE 2
MEAN MOTIVATION SCORES BY CHOICE GROUPS

Choice Group	N	Achievement Motivation		Affiliation Motivation	
		Mean	SD	Mean	SD
Success-subject only	18	8.67	2.99	3.44	1.71
Success-subject—friend	12	7.67	2.42	4.33	2.22
Friend—success-subject	28	6.46	2.65	5.00	2.31
Two friends	19	5.00	2.66	6.42	2.98
One friend	60	5.65	2.98	4.14	2.34

roughly proportioned to the marginal totals, but by far the largest number made F choices. This, too, is in accordance with expectation when the over-all tendency to F choices is considered. A chi-square test applied to this table yields a value of 61.71. The values of chi square for the individual motivation groups compared with the over-all frequencies are 21.87 for the High Achievement group, 11.70 for the Both High, 16.96 for the High Affiliation, and 11.18 for the Both Low. The significance levels are .001, .02, .01 and .05 respectively.

Table 2 presents the mean motivation scores of the Ss grouped according to choices made. The order of magnitude of both sets of means is as expected—achievement-motivation means decrease from the SS group to the FF and affiliation-motivation means increase. That the affiliation-motivation mean is not higher for the F group suggests that the F choice is a result of the characteristics of the situation rather than motivation of the individual. Analyses of variance of the two sets of motivation scores yielded *F* ratios of 14.52 ($p > .001$) for achievement and 4.74 ($p = .05$) for affiliation. The *t* tests computed on the differences between the individual means (not including the F group) revealed that although only one difference between adjacent means was significant, all four of the differences between means one-step-removed were significant at the .02 level or better and those between the extremes were significant beyond the .001 level.

These results suggest that scores on a test of motivation such as the one described provide a basis for predicting behavior in a choice situation. Apparently Ss with high achieve-

ment motivation are willing to overlook the fact that they don't particularly like another man if they think that man can help in the achievement of a goal. Conversely, those high in affiliation motivation don't care as much about goal attainment as interpersonal relations, and so choose their friends. In this situation, in fact, they generally chose both friends. The Ss with both motivations high were unable to make a choice and those with both low followed the line of least resistance.

Although the experiment was not designed with any analysis of the initial friendship ratings in mind, such an analysis was suggested by the great variability in the use of the 3, or dislike, category by the Ss. Apparently some Ss were unwilling to express dislike or hostility even when the recipient would remain unaware of it. This suggested the tentative hypothesis that these Ss would be high in affiliation motivation, possibly low in achievement, and would tend to choose friends rather than the SS in the choice situation. Those Ss who did use the dislike category might be expected to have lower affiliation motivation, possibly high achievement motivation, and be willing to reject friends and choose the SS.

As the fact that half the Ss made no 3 ratings on the friendship scale suggests, the distribution of friendship ratings was badly skewed. The Ss were therefore divided into two groups—those who made no dislike ratings and those who made one or more, and the number of each falling into the four motivation categories was determined. A trend in the predicted direction was apparent, with the High Achievement group having the greatest number of Ss who gave 3's and the fewest Ss who did not, and the High Affiliation group the reverse. A chi square over the whole table did not reach the .05 level but a comparison of the two extreme groups yielded a chi square of 3.85 ($p = .05$). The breakdown of the users of 3's vs. nonusers by choice-groups was also according to prediction with a chi square of 10.29, significant at the .05 level.

SUMMARY AND CONCLUSIONS

The present study was designed to test the hypothesis that the behavior of an individual making a choice between a work partner who was a competent nonfriend and one who was a less competent friend can be predicted from knowledge of the relative strength of achieve-

ment and affiliation motivation. The Ss first made friendship ratings and took a test designed to measure both achievement and affiliation motivation. Then they were divided into groups of four to work on an "important concept formation test." The groups were so constructed that three of the Ss, the critical Ss, had rated each other as liked and the fourth as not liked. They all worked individually on a sorting task at which the disliked S was made to succeed and the others to fail. Then all Ss wrote down either one first or a first and a second choice for a work partner.

Chi-square analyses of the four motivation combination groups indicated that, according to prediction:

1. The Ss high in achievement and low in affiliation motivation made significantly more single choices of the success-person.

2. The Ss high in affiliation and low in

achievement made significantly more choices of friend and the success-person.

3. The Ss high in both motivations made significantly more double choices involving both a friend and the success-person.

4. The Ss low in both showed no patterning.

In addition, the achievement-motivation mean scores showed a significant increase and the affiliation a significant decrease from Ss choosing the friends through Ss choosing both friend and success-person to Ss choosing the success-person only.

REFERENCES

1. FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.* 1955, 50, 232-236.
2. FRENCH, ELIZABETH G., & CHADWICK, IRENE. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, 52, 296-300.

Received October 3, 1955.

EFFECTS OF DIFFERENT CONDITIONS OF ACCEPTANCE UPON CONFORMITY TO GROUP NORMS¹

JAMES E. DITTES

Yale University

AND

HAROLD H. KELLEY

University of Minnesota

AMONG the variables influencing a person's conformity to a group's norms, two interrelated factors are likely to be important: (a) the extent to which he is attracted to or values his membership in the group; and (b) the extent to which he feels that other members are attracted to or value him. With respect to the first, it has been demonstrated that the more highly a person is attracted to a group, the more he conforms to the face-to-face pressures operating within it (1, 3) and the more he resists counternorm communications from outside it (6). The experiment reported here attempts to hold constant this first variable and to determine the effects on conformity of the second.

A relationship of *mutual dependence* is presupposed between the individual and group. Just as the individual is dependent on other members (a relationship emphasized in research on attraction to or valuation of the group), they also, to some degree, depend upon him—his skills, knowledge, and general ability to contribute to group life. Through their behavior they can be expected to communicate the value they place on him (to be referred to as their *acceptance* of him) and their desire to have him continue as a member. The present research varies this information concerning a person's acceptance in a group and investigates the effects on his conformity to its norms.

The effect of either variable on conformity is probably mediated by various intervening acquired motives. We would assume that in our culture, conformity to group standards is generally learned as a means of satisfying

many acquired motives, activation of which results, given appropriate conditions, in conformity behavior. The range of motives involved seems to include at least two fairly distinct clusters, one consisting of "approach" motives such as tendencies to help the group and to pattern one's self after admired persons, the other of anxiety-based motives such as desire to avoid social criticism and loss of membership.

It is probable that valuation of membership affects conformity through both kinds of motives, but that feelings of acceptance affect the person's sense of security about remaining in the group, which in turn affects primarily those motives based on anxiety. Assuming valuation constant, information communicated from fellow members that a person is little accepted by them increases his sense of insecurity, activating various acquired motives (to avoid social criticism, etc.), to which conformity behavior has been learned. Persons informed that they are well accepted in a group feel secure in their membership and have little anxiety-based motivation to conform.

In the present investigation, we are interested in two kinds of such information: (a) information about *how much* the person is accepted and (b) information about *how stable* the acceptance evaluations are. Either item would be expected to influence insecurity. Because of practical limitations, it was not possible to carry out a factorial design permitting evaluation of the separate effects and interaction of these two variables. It was decided that an investigation of their joint effects would provide a better initial test of the fruitfulness of our analysis than a study of either variable alone. Throughout the report, the label "acceptance" applies to both these components.

The hypothesis we propose to test is: Among persons who attach equal importance to their membership in a group, those who receive information that they are only minimally accepted by their colleagues and that this evaluation is subject to change, possibly

¹ This experiment was conducted as part of a research program on communication and opinion change operating at Yale University on a Rockefeller Foundation grant and directed by Carl I. Hovland. The authors gratefully acknowledge the excellent cooperation of the members of the Yale Class of 1958 who served as subjects. Members of the 1954-55 senior seminar on the psychology of the small group provided expert assistance in the conduct of the experiment. The authors are also indebted to Arthur R. Cohen, Carl I. Hovland, and Ben Willerman for their suggestions on the manuscript.

becoming worse, conform more than persons who receive information that they are highly accepted and that this situation is stable.

A similar hypothesis was tested by Kelley and Shapiro (5) who found it supported by correlational evidence that nonconformity was associated with high actual popularity or acceptance. But the hypothesis was not supported by the experimental manipulations of fictitious information about acceptance. It was suggested that this relationship was obscured because valuation of membership was not held constant. They found that information that one is poorly accepted affects not only his feelings of acceptance but also his subsequent valuation of the group. As our analysis suggests, a decline in valuation should decrease the "approach" motives which prompt conformity (e.g., desire to aid the group) thus offsetting any increase in other conformity-prompting motives resulting from the insecurity derived from feeling poorly accepted.

The present experiment was designed to avoid this obscuring effect in two ways: (a) It was attempted, with partial success, to keep valuation uniform by motivating all *Ss* so highly to participate in the group that their valuation of membership would remain constantly high throughout the experiment. (b) Several degrees of acceptance were introduced, instead of two, with the expectation that even if valuation varied significantly between the extreme conditions of acceptance, it might remain uniform between more similar conditions, particularly with the conditions of relatively high acceptance. In this case the hypothesis could still be tested by comparing the conformity among persons in those degrees of acceptance for which valuation remained relatively uniform.

METHOD

Subjects. One hundred and three volunteers from the Yale freshman class met in 18 five- and six-man groups. Members of each group had no prior acquaintance with one another.

*Incentives to motivate *Ss* to value and participate in their group, and instructions about possible rejection.* At the outset, each *S* agreed to work with his group in a contest against the other experimental groups. A cash prize and prestigious recognition were offered as awards for the group best in efficiency, smoothness of working together, and soundness of decisions. These instructions carried the strong implication that unanimity in group decisions was highly desirable.

It was then announced that to guarantee an effectively working group, members would have the option of eliminating from their group any person who appeared detrimental to its success. For this purpose, subsequent tasks were to be interrupted periodically while each member anonymously rated each other one on this question:

How desirable is it that this person be kept in the group?

- _____extremely desirable
- _____very desirable
- _____somewhat desirable
- _____not very desirable, but he should be kept in
- _____he should be rejected from the group

It was made to appear likely that some *Ss* would receive low ratings and that this would be unpleasant and penalizing for them: they would have to undergo the embarrassment of defending themselves in an open discussion of whether they should be rejected; rejection would carry the implication that the person was inadequate in his "social adaptability." (While the ratings were actually collected during the subsequent period, no *S* was ever discussed for rejection.)

Introduction of "delinquent gangs" problem and achievement of unanimous decision; interruptions to obtain ratings of others. In each group, an initial discussion concerned the relative worthiness of two gangs of juvenile delinquents, described in detailed simulated court records. A difference in the records insured that the same gang would be judged better in all groups, but this difference was slight enough that the decision could emerge only after considerable discussion. To heighten the tendency to conformity, the problem was described as being much like a jury's in that solution required the agreement of all. During the discussion, an observer kept a tally of each man's participation and whether his comments were for or against the gang ultimately favored by the group.

After general agreement was reached in free discussion, the norm was crystallized and registered by having the group rate the two gangs on several scales such that one gang was assigned labels such as "very deserving" and "fundamentally good"; and the other gang was characterized as "vicious," "malicious," etc. During this process *E* refused to record any rating until it was concurred in by every member. After this public agreement, *Ss* privately rated the gangs on different scales, to provide a measure of their initial acceptance of the group's characterization of the two gangs.

The discussion of the gangs was interrupted three times to permit the members to rate one another as to the desirability of their remaining in the group. Each time, *E* quickly thumbed through the ratings, pretended that no *S* had received low enough ratings to warrant discussing his possible rejection, and "discarded" the ratings into a wastebasket.

Introduction of different conditions of acceptance. At the end of the "gangs" discussion, *E* suggested a rest period. While *Ss* were relaxing, he casually inquired whether they'd be interested in seeing how they had been rated. Before anyone could object, *E* retrieved the slips from the wastebasket and distributed them. In this manner, each *S* was allowed to see privately what he thought to be the ratings made of him by the

other members. These were in fact fictitious ratings prepared in advance and substituted for the originals in the wastebasket. So that Ss would similarly interpret these ratings, *E* announced that he had noticed the average to be around "very desirable." In each group, one person found his ratings to be mostly higher than this average (*high* condition); two received ratings mostly at this average (*average* condition); two had ratings slightly below average (*low* condition); and one received ratings well below (*very low* condition). (In 5-man groups, only one of the intermediate conditions was represented by two persons.) Information about the probable stability of the ratings was provided by *E*'s statement, as though from experience with ratings in other groups, that the higher the rating the less likely it was to change; that low ratings were quite likely to change, possibly becoming higher or *even lower*. It was intended that as a result of this information, a person in the *high* condition would feel highly accepted, and that this was a stable situation. At the other extreme, a person in the *very low* condition was supposed to feel very little acceptance, and that his colleagues' evaluations might change, even becoming worse.

Before each session, the four conditions of acceptance were randomly assigned to positions around the discussion table. The 103 Ss were distributed among the conditions as follows: 18 in *high*, 33 in *average*, 34 in *low*, and 18 in *very low*. The intermediate conditions were assigned more Ss because greater variability of behavior was anticipated in them.

Additional information about the gangs, private judgments, and public discussion. After the fictitious ratings had been introduced and interpreted, the group proceeded to the second phase of the "gangs" problem. The purpose of this phase was to introduce pressure to deviate from the group consensus about the gangs and to determine each person's subsequent adherence to the norm, as indicated in his further private ratings of the gangs and in his opinions expressed in open discussion.

The pressure to deviate consisted of information suggesting that the unpreferred gang was superior to the previous evaluation, hinting at mitigating circumstances and desirable traits not clearly brought out in the original records. The Ss privately evaluated the additional information by checking agreement or disagreement with evaluative statements that accompanied it. The tendency to discredit or "explain away" the new information was assumed to represent conformity with the group's original norm (these scores are reported as Gang Index 1). The Ss also made additional ratings of the gangs after reading this information, one set anonymously and one set supposedly for public comparison (Gang Indices 2 and 3 respectively). To obtain an indication of each S's speed of responding to the contradictory information, the observer noted the order in which Ss finished making their evaluations and ratings.

Following the ratings, the group openly discussed the gangs for a short period during which were made observations comparable to those made during the initial discussion of the gangs.

Administration of number-judgment problem. A second situation for assessing conformity was next introduced, using a task requiring simple comparative judgments of numerosity—judging which of two squares contained

more dots, the same problem used by Kelley and Shapiro (5). The problem was presented to Ss working privately, in a series of eleven pairs of squares of decreasing difficulty, in which the correct answer was the same for the entire series. The Ss were to share ideas and try to improve their group score by exchanging written messages between successive judgments. In fact, their messages were not delivered, but were intercepted and replaced by a standard set of messages which led each S to believe that all the other members of his group had decided that a given square, *A*, was the correct answer for the series. To heighten the pressure to conform to this consensus, the rule was imposed that the group would score points on any one of the eleven successive judgments only when the group was unanimously correct.

Pressure to deviate from the consensus was introduced by having the later pairs in the series provide increasingly clear evidence that the consensus was in error, that square *B* was the correct answer.

On each pair, S reported his "private opinion" (and degree of confidence in it) and also a "public vote" to be tallied toward the group score. These were summed over the last ten judgments (the ones made following receipt of messages indicating the group consensus), yielding Dot Index 1 for the private opinions (weighted by confidence) and Dot Index 2 for the public votes. It will be noted that these conformity indices may reflect two aspects of conformity: (a) initial acceptance of the apparent consensus and (b) continued adherence to it in the face of the contradictory evidence. Actually, there is little variability in conformity on the early judgments (the general level being high), so these indices reflect primarily the second aspect.

Subsequent measures. (a) The Ss' perceptions of experimental conditions of acceptance were measured by a direct question immediately after they saw the ratings and at the end of the experiment, by recall of the ratings they had received. (b) Valuation of membership in the group: Immediately after receiving the ratings and also at the end of the experiment, Ss answered short questionnaires containing items measuring their positive motivation to remain in their particular group, e.g., desire to be invited to further meetings, liking to work with the other members. (c) The Ss' interpretations of conditions of acceptance: On a questionnaire given several months after the experiment, Ss indicated how likely they considered it to be that they might be rejected, and the freedom they felt about expressing opinions contrary to the group's judgment.

Termination of experimental session. At the end of each session, the purpose and procedures of the experiment were disclosed and the Ss were told that the ratings they had seen were fictitious. These disclosures were greeted with laughter, expressions of relief, and even a certain amount of disbelief. At no time did any S indicate that he had clearly doubted the genuineness of the ratings he had seen. Special care was taken to reinvolve as active and obviously accepted group members those Ss who had been in the conditions of *low* and *very low* acceptance. The Ss were requested to maintain secrecy during the remainder of the weeks scheduled for the experiment. Their excellent coopera-

tion in this provides one of several indications of the favorable attitudes with which they left the experiment.

RESULTS

Direct Effects of Experimental Conditions

Perceived acceptance in group. The fictitious ratings clearly produced the intended differences among the experimental conditions in amount of perceived acceptance. The question, "From the point of view of the group, how desirable is it that you, yourself, be kept in?" ("extremely desirable" scored as 5 and "I should be rejected from the group" scored as 1), yielded average scores of 4.7, 3.8, 3.3, and 2.9, respectively for the *high*, *average*, *low*, and *very low* conditions. Analysis of variance reveals that the between-condition variance is significant at beyond the .001 level of confidence. Furthermore, at the end of the experiment, Ss were able to recall their ratings accurately. The recalled ratings did not differ significantly from those actually received.

Perceived likelihood of being rejected. Several questions dealing with perceived likelihood of being asked to leave the group and with preoccupation about this possibility serve largely to differentiate the *very low* condition from the other three, the *very low* Ss viewing rejection as more likely. The *very low*'s differ significantly ($p < .01$) from each of the other conditions which do not differ significantly among themselves. These questions were asked only in the questionnaire given some months after the experiment so the results must be viewed with some reservations.

Valuation of membership. The efforts to keep all Ss highly attracted to the group were only partially successful. The *high* and *average* conditions showed the highest and approximately equal levels of attraction while the levels for the *low* and *very low* conditions were markedly lower, especially the latter. This effect was the same whether measured immediately after the ratings were distributed or at the end of the experiment. The total valuation scores yielded a between-condition variance which is significant at beyond the .001 level of confidence. Table 1 indicates the different valuation scores and the statistical significance of the differences between adjacent conditions.

Because valuation of membership was kept uniform only for *high* and *average* conditions,

TABLE 1
VALUATION OF GROUP MEMBERSHIP BY FOUR
EXPERIMENTAL CONDITIONS OF ACCEPTANCE

Condition of acceptance	High	Average	Low	Very Low
Average valuation of membership in group	40	39.5	37	35.5
Significance of difference between adjacent conditions	$p > .50$	$p < .0001$	$p < .0001$	

major interest in the subsequent results will be in comparisons of these two conditions. In these, we may expect there to be little interaction of valuation with our two main variables, acceptance and conformity.

Effects on Participation

The Ss in the *very low* condition reduced their participation in the discussion of the delinquent gangs by about 50 per cent after seeing the ratings ($p < .01$ for the change from the prior amount of participation). In contrast, *average* acceptance tended to result in increased participation, the difference between before and after being significant at the .09 level of confidence. The amount of participation after the ratings were received, expressed as a percentage of the amount of prior participation, is as follows: 93 per cent, 131 per cent, 118 per cent, and 56 per cent, respectively for *high*, *average*, *low*, and *very low*.

Effects on Conformity

Conformity in responding to questionnaires. Average values of the conformity indices derived from the questionnaires used with the two tasks are presented in Table 2, high positive scores indicating high conformity, negative scores low conformity. The gang-judgment indices were adjusted to take account of individual differences in initial conformity to the group decisions and intergroup differences in conformity behavior.² The three separate

² The scores on initial acceptance of the group decisions on the gangs were obtained before the experimental manipulations of acceptance and, as might be expected, an analysis of them reveals no differences approaching significance among the Ss who later found themselves in the different conditions. In order to eliminate these initial individual differences from the measures of subsequent conformity, each S's score on

TABLE 2
AVERAGE CONFORMITY SCORES FOR THE FOUR
EXPERIMENTAL CONDITIONS OF ACCEPTANCE*

Conformity Index	Experimental conditions of acceptance				<i>p</i> values for	
	High	Average	Low	Very Low	High vs. Average	Average vs. Low and Very Low
Gang judgments:						
1	-.22	.07	.10	-.09		
2	-.41	.32	-.10	.09	<.02	.10
3	-.31	.20	-.07	.04	<.10	
Over-all gang	-.92	.57	-.05	.02	<.05	
Dot judgments:						
1	39.0	43.4	33.2	36.4		.02
2	7.7	8.2	7.2	6.7		<.05
Over-all Conformity	2.5	3.2	2.5	2.6	.06	<.05

* Positive values indicate high conformity; negative values, low conformity. The significance of the difference between conditions on the over-all conformity index was tested with chi square, dividing those who were above their group mean on a majority of the five indices from the remainder. A *t* test was used to test the significance of the other differences noted.

gang indices were summed to give an over-all gang-judgment score. Also the five different measures were combined into an over-all conformity score by counting for each *S* the number of his scores on the five measures which were above the average of his experimental group. The average of these numbers for each experimental condition is presented in the last line of Table 2. Because the theoretical interest centers on the *high* and *average* conditions by virtue of their having the highest and approximately equal levels of valuation of membership, *p* values are given in Table 2 for differences between them. Since the two

initial acceptance was subtracted from each of the three later measures of opinion. Analysis of variance of each of these difference scores showed that the 18 experimental groups differed significantly in mean level of conformity and also in variability. These effects presumably reflect slightly different definitions of the norm and differing pressures to conform which developed during different experimental sessions. To eliminate these differences, on each measure of the gang judgment each *S* was given a standard score representing the amount of his conformity with reference to the mean and standard deviation of his particular experimental group. As noted in the text, the dot-judgment scores, based only on behavior occurring after the experimental manipulations, reflect both initial acceptance of the norms and resistance to contrary evidence, though largely the latter. It was not necessary to transform them to standard scores because they did not vary significantly among experimental sessions with regard either to means or variances.

low conditions show the same tendency to be lower than the *average*, *p* values are also presented for the differences between their combined mean and that of the *average*.

In general, *Ss* receiving ratings of *average* acceptability tended to show more conformity behavior than did *Ss* receiving ratings of *high* acceptability. The results with *low* and *very low* conditions are less uniform, possibly for reasons connected with their rather low valuation of membership. They showed less conformity than the *average* and more than the *high* condition on gang judgments, and less than either of the other two on dot judgments.

The three indices of conformity on gang judgments were found to be significantly intercorrelated, as were the two indices on dot judgments. Significant correlations between gang and dot measures were found only for the *high* and *average* conditions and these had values of around .50. Despite this low degree of consistency between the two types of conformity behavior, Table 2 shows that the difference in conformity between *high* and *average* conditions of acceptance is consistent over the two somewhat different tasks.

A less direct indication of conformity gives results supporting those of Table 2. An observer noted the order in which the *Ss* at each session completed the task of evaluating the additional information that suggested that the group's decision on the gangs was inaccurate. It might be assumed that persons oriented most strongly toward conforming would give less attention to the new information and make quick judgments simply on the basis of the previously established norm; and that persons less motivated to conform would give more conscientious attention to the information and hence complete the task more slowly. The average rank order of finishing the task is as follows: 4.2, 2.7, 3.8, and 3.1, for *high*, *average*, *low*, and *very low*, respectively. The condition of *average* acceptance was fastest, indicating, on the basis of the above assumptions, most conformity; the *high* condition was slowest indicating least conformity. The difference between these two is significant at the .05 level of confidence, using a chi-square analysis that divided *Ss* into slowest and fastest halves.

Conformity in open discussion. In Table 2, the measures intended to tap private opinions (Gang Index 1 and 2; Dot Index 1) and those

supposed to reflect public expression (Gang Index 3 and Dot Index 2) reveal essentially similar trends over the four conditions of acceptance. However, as compared with the quasi-public nature of the latter measures, analysis of actual public expressions of opinion shows a somewhat different pattern of conformity from that noted in Table 2. (This finding suggests that the questionnaire results mainly reflect private conformity.) In the discussion of the gangs after the additional information had been read, there was a general decrease in the proportion of comments that expressed conformity to the group's earlier preference for one gang over the other. (Before the information, 88 per cent of the comments involving opinion were favorable to the group norm; afterward, only 49 per cent were so.) This decrease was less for the *very low* condition than for the others; i.e., in their actual public remarks, the *very lows* conformed more closely to the original norm than any of the other conditions. The scores³ for the *high*, *average*, *low*, and *very low* conditions, respectively, were: $-.55$, $-.47$, $-.52$, and $-.23$, with the greater negative number indicating the greater deviation from the norm. Chi-square analysis, dividing *Ss* at the over-all median into high and low changers, shows that the difference between *very low* and the other conditions taken together is significant at the .08 level. The difference between any other condition and the rest does not approach significance.

This result receives partial confirmation from responses to a question about how free the *S* would have felt to express an opinion contrary to the group opinion. *Very low Ss* report more hesitancy in this respect than do *High Ss* ($p = .07$) and the other conditions report intermediate degrees.

DISCUSSION

To simplify the reader's task of assimilating the results and trends reported above, they are summarized verbally in Table 3. We now consider the interpretation of these findings. In several respects, they are consistent with

³ The score for each *S* represents the proportion of his total remarks in the later discussion which were pronorm, minus the proportion of his total remarks in the earlier discussion which were pronorm, plus the proportion of antinorm remarks in the earlier discussion, minus the proportion of antinorm remarks in the later discussion.

the hypothesized relationship between acceptance and conformity. However, the total pattern of results suggests that the relationship depends on more complex intervening variables than had been anticipated.

One set of data consistent with the hypothesis is provided by the comparison of the *high* and *average* conditions. Since the relation between acceptance and conformity is likely to be obscured if valuation varies, as previously discussed, the hypothesis can be clearly tested only by comparing conditions having fairly similar levels of valuation—in this experiment, the *high* and *average* conditions. Between these two, higher conformity behavior was exhibited by the condition of lower and less stable acceptance. The available evidence does not indicate, however, any difference between *high* and *average Ss* in feelings of insecurity about their membership. If these groups actually do not differ in this respect, the higher conformity of the *average Ss* must be explained on some other basis than the one underlying our hypothesis: that they would be more motivated to avoid rejection. Their total pattern of behavior—high conformity on the questionnaires, rapid reactions to counternorm information, and heightened participation in the discussion after

TABLE 3
DIGEST OF THE RESULTS OF THE FOUR EXPERIMENTAL
CONDITIONS OF ACCEPTANCE

Variable	Experimental condition of acceptance			
	High	Average	Low	Very Low
Perceived acceptance	High	Average	Low	Very Low
Perceived likelihood of being rejected				Higher than any other condition
Valuation of membership	Moderate	Moderate	Low	Very Low
Participation in group discussion		Highest		Lowest
Speed of reacting to counternorm information		Faster than any other condition		
Conformity in questionnaire responses (private?)	Low	High	Low	Low
Conformity in public statements				Higher than any other condition
Felt freedom to express deviant opinion	High-est			Lowest

learning of their acceptance—is perhaps more suggestive of a desire to improve their social standing. The *S* in the *average* condition places considerable value upon the group, but at the same time, is not completely accepted in it. Since *E*'s comments indicate there is some possibility that his acceptance status may change, we might expect him to try to attain complete acceptance. Eager participation and uncritical conformity to the norms would presumably facilitate this improvement in status. That this conformity extends even to the expression of private opinions on the questionnaire is understandable if we consider that there may be a general tendency for upward-mobility-oriented persons to identify with and take over the values of higher status individuals (7).

The behavior exhibited by the *very lows* is much more consistent with our initial notion of anxiety-motivated conformity. In general, and in accord with prior findings (4, 5), group membership loses its attractiveness for persons little valued by their colleagues.⁴ Along with this loss goes a decline in tendency to accept the group's norm, at least in quasi-private settings. In the extreme case, however, where acceptance is so low that actual rejection is presumably an imminent possibility, anxiety about rejection is especially high and the result seems to be a pattern of guarded public behavior, i.e., the *very lows* withdrew from open participation in the discussion and showed the highest amount of conformity in the opinions they did publicly voice. Avoidance of rejection, which is presumably motivated by penalties particularly attached to rejection (embarrassment, suffering failure in front of the *E*), is of course a variable different from the positive motivation to participate in the group; so anxiety over rejection may have been high for the *very lows*, even though they had little positive motivation to stay in the group. (The questionnaire some months after the experiment indicated that the desire to be kept in the group and not be rejected had been as high for the *very lows* as for the other condi-

tions, while at the same time, their estimate of the likelihood of this event had been higher.) Their public conformity, then, may reflect a relatively high motivation to avoid rejection. That they exhibit a high public conformity but little private acceptance is consistent with Festinger's hypothesis that this particular pattern of conformity results when the attractiveness of group membership is low, but people are constrained to stay in the group by external threats or barriers against leaving (2).

In brief, our results suggest two processes linking conditions of acceptance to conformity. For persons who value a group, are less than completely accepted in it, but have some possibility of achieving complete acceptance, conformity facilitates such a gain in acceptance. Their conformity is unquestioning and extends to private opinions as well as to public behavior. On the other hand, persons who are on the brink of unwelcome rejection manifest conformity only at the public level, presumably as a means of forestalling such rejection.

The major variable in this experiment—security and degree of acceptance—is, from several points of view, an aspect of what is commonly called “status.” We might therefore expect our results to shed some light on the problem of the relationship between status and conformity. In general, we doubt that this relationship is ever a simple one. Our results and interpretations suggest that to predict conformity from status, clarity is first needed as to the conditions under which conformity is to be observed, whether they involve surveillance by other members or relative privacy. Then, it must be asked whether the observed differences in status are associated with differences in (a) valuation of membership, (b) security in membership, and (c) motivation to improve one's acceptance. Under some conditions, as when low status is associated with relatively high insecurity about membership, it seems likely that status is inversely related to public conformity. But if there is a marked difference between high and low status persons in their valuation of membership, as a result of greater privileges and satisfactions being associated with higher status, status and conformity may be correlated positively.

Undoubtedly, a number of other factors also affect this relationship. For example, it is easy to imagine instances where special motives directed toward conformity would be opera-

⁴ This raises interesting theoretical questions which cannot be discussed here: Is acceptance by others one of the prerequisites for desiring membership in a group? Why does valuation not decline with the change from *high* to *average* conditions, but drops sharply for the lower degrees of acceptance? Is this a level-of-aspiration effect, in which the motivation to stay in the group declines only as the possibility of success in this effort drops below some critical value?

tive for high status persons, but not for low status members. One such case would be that where the opinions of a high status member have carried inordinate weight in the setting of group standards so that they happen to coincide with his private preferences. What appears as conformity in his behavior may actually be motivated by private considerations that preceded the acceptance of his behavior as the norm. Another instance may be when special conformity demands are made of high status persons, for example, because of the special symbolic value their conformity may possess.

These considerations indicate that many factors affect the relation between status and conformity. In some instances, the relationship may be direct; in other, inverse; and in still others, perhaps, nonexistent. We believe the specific situations and results will be most intelligible when analyzed in terms of factors such as those suggested above.

SUMMARY

In experimental groups of Ss, two different norms were developed: one concerning a social value judgment; the other, a simple perceptual judgment. The Ss were experimentally made to feel different degrees of being accepted by the other members and were then given opportunities and incentives to deviate from the norms. Subsequent conformity, participation, and attitudes toward the group were studied in relation to the different conditions of acceptance.

The results point to two contrasting patterns of conformity evoked by different conditions of acceptance. The first appears to consist of a high degree of genuine adherence to the norms, as indicated by unquestioning conformity extending even to conditions of privacy, and a higher-than-average motivation to participate in group discussion. This pattern appeared for the experimental condition in which subjects enjoyed somewhat less than

complete acceptance but probably saw the possibility of gaining this status. It is interpreted as based on strong positive attachment to the group and motivation to improve one's status therein.

The second pattern is marked by high conformity only under public conditions. It occurred for the experimental condition of lowest acceptance, in which Ss saw total rejection as being a likely possibility. The interpretation is made that although Ss in this condition have lost much of their positive motivation to conform to group standards (indicated by their very low valuation of membership and reflected in their low private conformity), they may nevertheless be concerned about the negative consequences accompanying rejection (at least, under the conditions represented in this experiment). Public conformity is seen as a way of forestalling this unpleasant eventuality.

REFERENCES

1. BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9-23.
2. FESTINGER, L. An analysis of compliant behavior. In M. Sherif and M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 232-255.
3. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.
4. JACKSON, J. Analysis of interpersonal relations in a formal organization. Unpublished doctor's dissertation, Univer. of Michigan, 1952. Reported in D. P. Cartwright and A. F. Zander (Eds.), *Group dynamics*. Evanston: Row, Peterson, 1953. P. 424.
5. KELLEY, H. H., & SHAPIRO, M. M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *Amer. sociol. Rev.*, 1954, **19**, 667-677.
6. KELLEY, H. H., & VOLKART, E. H. The resistance to change of group-anchored attitudes. *Amer. sociol. Rev.*, 1952, **17**, 453-465.
7. LEWIN, K. Self-hatred among Jews. In *Resolving social conflicts*. New York: Harper, 1948. Pp. 186-200.

Received October 3, 1955.

VERBAL REINFORCEMENT AND INTERVIEWER BIAS¹

DONALD C. HILDUM AND ROGER W. BROWN

Harvard University

IN SEVERAL recent studies verbal behavior has been manipulated by means of selective reinforcement. These studies have very disturbing implications for the clinical and public-opinion interview. In the initial experiment Greenspoon (2) asked college students to voice nouns *ad libitum*. For one group of Ss, E murmured "Mm-hmm" whenever a plural noun was produced, while for a control group he said nothing at all. The reinforced Ss used more plural nouns than the control Ss. A somewhat different technique was developed by Taffel (4) and also employed by Cohen *et al.* (1). They required S to form a sentence when given a verb and a choice of one of six personal pronouns to be used in starting the sentence. It proved to be possible to affect the frequency with which S used the various pronouns by saying "Good" in a flat unemotional tone after sentences employing the desired pronoun subject. Finally, Greenspoon has reported (3) the successful use of two nonverbal stimuli (a red light and a 190-c.p.s. tone) to increase the frequency of plural and also of nonplural responses. In all of these studies S said that he had not been conscious of any connection between his own behavior and the reactions of E. This is learning without awareness.

In these experiments reinforcement seems to act like the moon on the tides with inevitable and uncontingent effect. For those interested in such things there appears a glittering new prospect for human manipulation; for others only the quiet pleasure to be found in any new proof of human stupidity. The verbal-reinforcement data will confirm many in their suspicion of the methods of clinical and social psychology. The therapist who believes in the importance of the Oedipus complex could elicit Oedipal content by means of selective reinforcement. Perhaps a patient could even be brought to an appearance of mental health through the encouragement of "healthy" utterances. Is not "Mm-hmm" the very hallmark of a therapeutic school? In an opinion interview

the S might be infected with the opinions of the interviewer by means of patterned reinforcement. Perhaps verbal reinforcement is a mechanism of interviewer bias.

It can be answered to these alarms that neither interviewers nor therapists wish to influence their Ss and consequently they would not use verbal reinforcement. There is no comfort in this answer. Anyone who has worked with verbal reinforcement knows that our "Mm-hmm" and "Good" are ordinarily unconscious and automatic. They can only be controlled after some practice. It follows that an interviewer might influence his S in all innocence and we know from the experiments that an S might unwittingly accept the influence. Here is powerful social influence operating outside the awareness of everybody concerned.

Two important differences between the laboratory studies and actual interviews should however be noted: (a) Greenspoon and the other experimenters have taken as their response units particular words or word types. Interviewers are not ordinarily pleased either by plural nouns or by the use of the first person singular. If they are moved to agree or disagree it is likely to be with reference to a line of thought or an expressed attitude. In short, interviewers would be likely to reinforce content categories rather than specific verbalizations. (b) In a face-to-face conversation there are many ways of communicating agreement or disagreement—smiles, nods, averted eyes. In the laboratory studies of verbal reinforcement E sometimes placed himself behind the S and sometimes faced him but endeavored to avoid giving visual cues. These efforts at isolating the independent variable may not have succeeded. In any case they must have created a rather strange social atmosphere unlike the usual interviewing situation.

We undertook an experiment intended to correct these two laboratory artificialities. The experiment took the form of an attitude survey conducted by telephone. The interviewer asked questions concerning the Harvard

¹ This study was supported by the Laboratory of Social Relations, Harvard University.

philosophy of General Education. He revealed a bias for or against the philosophy by his reactions to the answers of the interviewee. In effect, he played the role of one who favored or of one who opposed General Education. However, the role-play was limited to the single matter of verbal reinforcement. The *S* could discover *E*'s bias from his first few reactions and might then provide more of the kind of content that was reinforced. Administering the questionnaire by telephone created a situation in which vocal reaction was the only kind possible and so the independent variable was isolated in a natural way.

METHOD

The questionnaire. In order to maximize the likelihood of obtaining a verbal reinforcement effect, a topic (General Education) was selected on which *S*s were not expected to have very strong opinions and so were expected to be open to influence. Also, though *E* might be expected to have an opinion, there was no obvious stereotype to tell *S* what that opinion should be. In the questionnaire there were 15 questions, with 4 possible responses to each: Agree strongly, agree slightly, disagree slightly, disagree strongly. The statements were worded so that agreement with some statements represented an attitude favorable to General Education while agreement with other statements represented an unfavorable attitude. Consequently the interviewer reinforced an attitude rather than a particular response. To make sure that the favorable or unfavorable nature of each statement was clear, we originally wrote 40 statements intended to sound unlike but really covering much the same ground. From these we selected 15 which on pretest proved to constitute a maximally redundant subset. It is generally possible to predict from the answer to one of these questions the answers to all the others. In addition the questionnaire was prefaced with a brief statement defining the philosophy of General Education so as to make sure that the *S* would understand which statements favored the philosophy and which were opposed.

Subjects. Forty male students, graduate and undergraduate, were sampled from those in the Harvard Summer School who listed Cambridge telephone numbers and home addresses in the United States. Ten *S*s were assigned at random to each of four experimental groups.

Procedure. The same *E* conducted all the interviews. He identified himself as a member of the Social Relations Department and said that the survey was designed to compare summer-school opinion with that of the rest of the student body. He then administered the questionnaire.

For two groups the reinforcement was "Mm-hmm," pronounced after pro General Education answers for one group, and after anti answers for the other group. For the remaining two groups the reinforcement was "Good," again with one group rewarded for favorable responses and the other for unfavorable responses. Both

reactions were fed back immediately following an approved answer in a neutrally-toned, rising inflection. The *E* (a trained linguist) carefully equated the intonation used in reading questions and responding to *S*s in all groups.

At the close of the interview *S* was asked to guess *E*'s opinion on General Education and also to say whether he thought there was any bias in the questionnaire or in its administration. If *S* thought there was bias or felt that he knew *E*'s opinion he was then asked whether his answers were influenced by these conditions.

RESULTS

The reinforcement effect. Each answer was scored from 1 to 4: 1 when an answer was strongly in favor of General Education, 2 when slightly favorable, 3 when slightly unfavorable, and 4 when strongly unfavorable. Refusals to answer (12 of 600 responses) were scored 2.5. The total score for an *S* had a possible range from 15 to 60. The actual range for all *S*s was from 15 to 39.5. The *S*s tended to favor the General Education philosophy.

In Table 1 appear the mean attitude scores for the four experimental groups and in Table 2 are the *t* scores and probabilities for the crucial comparisons among these means. Reinforcement with "Good" was effective while reinforcement with "Mm-hmm" was not effective. In fact, with "Mm-hmm" the obtained difference was in the opposite direction to that predicted from the reinforcement scheme.

TABLE 1
MEAN ATTITUDE SCORES FOR FOUR
EXPERIMENTAL GROUPS

Group	Mean	SD
"Good"-pro	24.95	5.10
"Good"-anti	31.75	3.46
"Mm-hmm"-pro	29.8	3.96
"Mm-hmm"-anti	27.1	4.06

TABLE 2
t SCORES AND PROBABILITIES FOR CRUCIAL
COMPARISONS AMONG THE FOUR
EXPERIMENTAL GROUPS

Comparison*	<i>t</i> score	<i>p</i>
"Good"-pro vs. "Good"-anti	3.31	.01
"Mm-hmm"-anti vs. "Mm-hmm"-pro	1.43	.20
"Good"-pro vs. "Mm-hmm"-pro	2.25	.05
"Mm-hmm"-anti vs. "Good"-anti	2.62	.02

* First member of comparison has score more favorable to General Education.

With the same pattern of reinforcement "Good" was more effective than "Mm-hmm."

Awareness of reinforcement. Eight of 20 Ss noticed that the interviewer said "Good" but only 1 of these assumed that it indicated approval. Only 1 of 20 Ss noticed that *E* said "Mm-hmm." He thought it might have meant approval. All Ss rejected the notion that their answers had been influenced by the interviewer's reactions.

Of 24 Ss who were willing to guess at *E*'s opinion all but one said that *E* favored General Education. Apparently they based this guess on the assumption that someone surveying opinions on General Education would himself favor the principle, rather than on the pattern of reinforcement.

Replication of the experiment. The essential design was repeated by members of an undergraduate tutorial group with 25 Ss. The results were like those of the initial study. The mean "Good"-pro score was lower than the mean "Good"-anti score with $p < .05$, while the mean scores for the two groups reinforced with "Mm-hmm" were not significantly different. The mean "Good"-pro score was lower than the mean "Mm-hmm"-pro score with $p < .01$. The mean "Good"-anti and "Mm-hmm"-anti scores did not differ significantly. This study involved new interviewers and Ss enrolled for the regular academic year at Harvard rather than in summer school.

DISCUSSION

In the present interview situation "Good" is a reinforcer while "Mm-hmm" is not. This result conflicts with Greenspoon's finding that "Mm-hmm" affects the frequency of plural nouns. If conscious learning were involved, an explanation for the discrepancy would be available. When an *S* free-associates aloud for 25 minutes, he may be expected to search rather desperately for some indication from *E* as to the purpose of the task and the proper direction to take. In these circumstances he will notice "Mm-hmm" or even a red light or a tone of 190 c.p.s. and be guided by them. When *S* is asked to give his opinion of a set of statements, he is engaged in a task that makes sense as presented, and consequently he will not attend so closely to *E*'s reactions. In this situation the muttered "Mm-hmm" should have no effect. The difficulty is, of course, that

Ss in our experiment and in Greenspoon's nearly all said they had been quite unaware of *E*'s "Mm-hmm." In other words the two sets of data give no indication of differential attention to *E*'s reactions and so do not support our explanation. Still, the explanation may be correct and our measures of awareness insufficiently sensitive.

There are many questions one might ask to test *S*'s awareness. Does he know the purpose of the experiment? Did he notice anything about *E*'s behavior? Did he notice that *E* said "Mm-hmm?" Does he realize that these "Mm-hmms" have influenced his behavior? All of the studies so far find that Ss usually do not attain to this last level of awareness. It is not clear from the experimental descriptions which other questions were asked nor what the answers were. It is possible, therefore, that reinforcers are more effective on the higher levels of awareness than on the lower levels and that these levels have not been adequately distinguished in the studies reported.

We did ask our Ss several questions before inquiring whether they knew about *E*'s influence on their behavior. Eight Ss were aware that *E* had said "Good" while only one *S* noticed "Mm-hmm." Evidently "Good" is more "visible" in the interviewer's role than is "Mm-hmm." Probably this is because saying "Good" very nearly violates the prescribed nondirective character of that role. This is a difference of awareness that may help to account for the effectiveness in our study of "Good" and the ineffectiveness of "Mm-hmm." We also asked our Ss to guess at *E*'s opinion of General Education. We thought it quite possible that while *S* usually fails to notice *E*'s responses he might react to them in forming a conception of *E*'s opinions and this conception might be verbalizable when its behavioral sources were not. This measure disappointed us in the present case, in that all but one *S* thought *E* favored General Education. However, there is a tendency for the Ss who were reinforced for anti-General Education answers to attribute to *E* a somewhat less favorable attitude than do the Ss reinforced for pro-responses. Furthermore the Ss whose conceptions of *E*'s attitude show more sensitivity to the pattern of reinforcement received are also the Ss whose own attitude scores seem to have been most influenced by *E*'s reactions.

While all of these trends fall short of significance in the present study, they do open the possibility that levels of awareness will help to account for differences in the effectiveness of selectively interpolated experimenter reactions.

There is a further point of contrast between our results and those of Greenspoon. Since a light and a tone have been used as reinforcers it would seem that semantics plays no necessary part in this phenomenon. Yet we find "Good" effective and "Mm-hmm" ineffective, and we are inclined to think that this may be partly due to the fact that "Good" has a more clearly favorable meaning than "Mm-hmm." The meaning of "Mm-hmm" is altogether dependent on the intonation pattern. As a sequence of segmental phonemes, as a printed word, its meaning is ambiguous. It may be a neutral indication that one is listening or it may even be questioning or disapproving. "Good" is somewhat less dependent on intonation. It is more reliably a favorable sign. This semantic difference may help to account for the superior effectiveness of "Good."

Our experimental procedure has brought the verbal reinforcement experiment nearer the interview situation. It may eventually prove to be necessary to train interviewers to control their specific reactions to the content received from an informant. Some of these reactions (smiles, averted eyes, etc.) may not be susceptible of control. It may be wiser, in a study of public opinion, to use interviewers of opposed bias, letting their reinforcing reactions operate freely but hoping the effects will cancel one another in the total sample. Pollsters might even learn much about opinion stability by having interviewers deliberately take different sides for different Ss.

In any case, there is much to be learned about the reinforcement of verbal behavior before worrying unduly about its effects in the interview. It has not yet been shown, for instance, that a reaction like "Mm-hmm" is

selectively used to express approval in ordinary conversations or interviews. The reaction *can* be selectively interpolated by an experimenter but *is* it selectively interpolated outside the laboratory? Can one predict the naive speaker's use of "Mm-hmm" from knowledge of his attitudes? Even if this be possible, it does not follow that the usage of either the clinician or the opinion interviewer is similarly predictable. Their efforts to play a nondirective role may well change the ordinary usage of "Mm-hmm." Actual interviewer behavior needs to be studied before deciding on the importance of the studies of verbal reinforcement for the interview.

SUMMARY

A questionnaire was administered by telephone and the interviewer attempted to influence his Ss through the selective interpolation of two reactions—"Good" and "Mm-hmm." "Good" proved to bias the results obtained while "Mm-hmm" did not. The study was repeated and the same result obtained. These results were compared with those obtained by other experiments. The implications of the verbal reinforcement phenomenon for the clinical and opinion interview are discussed.

REFERENCES

1. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106-110.
2. GREENSPOON, J. The effect of verbal and non-verbal stimuli on the frequency of members of two verbal response classes. Unpublished doctor's dissertation, Indiana Univer., 1951.
3. GREENSPOON, J. The effect of two nonverbal stimuli on the frequency of members of two verbal response classes. *Amer. Psychologist*, 1954, **9**, 384. (Abstract)
4. TAFTEL, C. Conditioning of verbal behavior in an institutionalized population and its relation to "anxiety level." Unpublished doctor's dissertation, Indiana Univer., 1952.

Received October 13, 1955.

THE GENERALITY OF COGNITIVE COMPLEXITY IN THE PERCEPTION OF PEOPLE AND INKBLOTS

JAMES BIERI AND EDWARD BLACKER

*Harvard University*¹

THEORIES of behavior that use perceptual or cognitive constructs have found it necessary to postulate some organizing or schematizing process which is held responsible for the active interpretation and representation of external events to the organism. The ego of Freudian theory is considered to mediate between the environment and the organism, while Lewin speaks of the "functional firmness of the boundaries between individual and environment" in discussing the impact of the environment upon the individual (11, p. 107). Tolman (14) discusses the conditions under which the "cognitive map" of the organism operates as an intervening variable in behavior. Bartlett (1) utilizes the concept of "schema" to refer to the organization of previous experience which effects behavior in a current situation. Piaget (12) discusses the developmental aspects of the assimilation and differentiation of the environment. In a recent formulation, Kelly (9) has utilized the notion of a "personal construct system" to help explain the individual's behavior in actively perceiving and responding to the meaningful environment. All such formulations imply two things about the role of perception in behavior. First, perception is an active process involving a transformation of sensory data into a conceptual scheme consistent with the previous learning and experience of the individual. Secondly, the conceptual scheme is structured differently from one individual to another, and an understanding of these *structural* differences is of value in predicting the behavior of the individual. Thus, in Freudian theory the ego has certain reality-testing functions which are affected by its stage of development and its characteristic defenses. Lewin speaks of the degree of differentiation of the inner personal regions of the life space and Tolman discusses the effects of broad and narrow cognitive maps on behavior. If we assume that every individual has a cognitive system or scheme for construing his environment, there remains the problem

of how we characterize the structure of this system so as to obtain meaningful predictions about behavior.

In an earlier paper (3), the position was taken that one way of approaching the problem of cognitive structure was to posit a dimension of complexity-simplicity. That is, an inference can be drawn about the relative complexity of an individual's cognitive system from the measured complexity of his perceptions of external events. Operationally, complexity is measured by the variability of responses given to a finite realm of stimuli. Essentially, we are asking, in how many different ways can a person perceive a certain set of events? The assumption is made that the more differentiated the system is relative to these events, i.e., the more alternative *perceptions* of an event are available, then the more alternative *behaviors* will be available in that situation.

An important question arises as to the generality of an individual's cognitive system from situation to situation. It seems evident that an individual with a developmental pattern of experiences varying from very intensive in some situations to little experience in others will have learned to approach these different situations with varying levels of cognitive complexity. However, it seems unwise to postulate a capricious specificity to the individual's cognitive behavior from situation to situation. Rather, it is assumed that relatively consistent, enduring modes of cognitive schematization will characterize the individual's behavior across situations. Genetically, we assume that as the individual's cognitive system develops in one realm of experience, it will tend to generalize to some extent to new realms of experience subsequently encountered by the individual. Vernon (15) has recently discussed the functions of "schemata" in different types of experimental perceptual situations. Vernon and others have criticized experimental studies of form qualities, constancy, and motivational factors as perhaps relying too much on transitory,

¹ This study was facilitated by a grant from the Laboratory of Social Relations, Harvard University.

artificial states rather than upon states and situations more readily encountered in everyday life.

In line with this criticism, and in view of the importance of the perception of people for personality theory, we would emphasize the primary importance of the individual's cognitive system as it is organized about the perception of people in his life space. Thus, among the first objects in the environment with which the infant has intimate contact are people. As most personality theories stress, the period of infancy and early childhood is one in which important, perhaps basic, perceptions and differentiations are developed relative to the future course of the individual's interpersonal experience. Might it not be then, that the structure of the cognitive system the child learns in relation to his perception of people is basic in determining the structure of the system relative to other stimuli and events in his life? Without being concerned now about the *causal* sequence involved, we can formulate the fundamental problem to be investigated in the present study: Can there be established a degree of generality between the complexity of the individual's cognitive system in the perception of people and in the perception of events that fall outside the immediate realm of interpersonal experience?

The measure of cognitive complexity in the perception of people that is used in this study has evolved from a series of attempts (3, 9) to express objectively the degree of differentiation an individual possesses in his perception of others. This measure will be described in detail in the section on method. The measure of complexity of cognitive behavior using non-human stimuli was derived from the behavior elicited on a modified Rorschach-test situation. Rorschach inkblots were used not only because of the importance and relevance of this instrument to clinical practice and research, but also because it affords two kinds of behavior which can be analyzed, namely, the *determinants* utilized in giving a response and the *content* of the response itself. In a manner to be discussed below, both the determinants given by an individual and the content given by an individual can be conceptualized as to their relative complexity. Again, it should be mentioned that the basic operation for measurement of cognitive complexity is essentially

response variability to a prescribed realm of stimuli.

With these considerations in mind, we are now in a position to state the hypotheses to be tested:

I. Significant positive relationships exist between the complexity of subjects' perceptions of people and the complexity of their perceptions of inkblots relative to the determinants used.

II. Significant positive relationships exist between the complexity of subjects' perceptions of people and the complexity of their perceptions of inkblots relative to the content elicited.

METHOD

Subjects (Ss) in this study consisted of 40 male university undergraduates who were paid to serve in psychological experiments.

Complexity Measures

Perception of people. The Repertory Test (Rep Test) devised by Kelly (9) for evoking "personal constructs" was used to obtain the perceptions of others. This technique consists of presenting *S* with the names of three persons known to him and asking *S* to sort the three in such a manner that two are perceived as alike or similar in some important personal way and different from the third in this respect. After *S* states the similarity of the two sorted together, he is asked the opposite of this perceived characteristic. Thus, *S* may sort two persons as similar to each other and different from the third because the two are "sincere," and the opposite of this for *S* might be "insincere." After the initial perception and its opposite are obtained, *S* is presented with three more persons and asked to sort them in an analogous fashion. The names of the persons sorted have been placed on cards by *S* according to a standard list of "role titles" supplied by the experimenter (*E*). In the present study, the following role titles were used on the Rep Test:

- a. Yourself
- b. Your brother closest to you in age (or person most like a brother)
- c. Your closest girl friend
- d. The most successful person whom you know personally
- e. Someone you know personally whom you admire
- f. Someone you know personally whom you would like to help or for whom you feel sorry

These role titles were selected from a larger group of role descriptions used in previous research as offering maximum discrimination in terms of perceptions elicited. The six persons were sorted by *S* in all possible combinations of three such that no three names appeared together in a sort more than once. In all, 20 sorts were thus obtained, with a perception of similarity and its opposite being obtained on each sort. Essentially, then, *S* was confronted with a task in which he was asked to perceive a fixed number of stimuli in dif-

ferent combinations. The complexity of *S*'s perceptions of others was measured in terms of the number of different perceptions (NDP) *S* gave in the 20 sorts. If, in giving a perceived similarity and difference on a sort, *S* gave the same verbal similarity or difference which he had given on a previous sort, it was scored as a repeated perception. It should be noted that a stringent criterion of response repetition was used, i.e., *S* had to use the same word or words in his perceived similarity or difference as used previously. Theoretically, *S* could receive a maximum complexity or NDP score of 20 or a minimum NDP score of one.

The NDP scoring was done by one of the authors. The reliability of this scoring was obtained by selecting ten protocols at random and having a judge score these records independently according to the criteria outlined above. The interscorer reliability coefficient for these ten cases was .92. In addition, an estimate of the stability of the NDP measure was obtained by repeating the sorting procedure with 19 *S*s after an interval of two weeks. The test-retest reliability coefficient obtained was .82. The magnitude of this latter coefficient indicates an adequate degree of stability in this measure of the complexity of perceptions of others.

Perception of inkblots. Approximately two to three weeks after the Rep Test was administered, each *S* was given individually the modified Rorschach procedure so that every *S* would have a constant series of stimuli to perceive, a series that would be repeated so as to afford opportunity for variable interpretations to be evoked to the same stimuli. Thus, both the Rorschach situation and the Rep Test contain the crucial features necessary for our complexity measure.

The Rorschach modification used has been described in detail in another paper (4). It consists of ten Rorschach large details (*D*), one from each of the ten cards. All other portions of the blot are blocked out by means of a template which exposes only the desired blot portion. Five of the blots are chromatic and five are achromatic. The particular blot details used were selected because of their judged ability to evoke a variety of responses, both in terms of content and in terms of determinants. Beck's notation (2) for the ten details thus selected, in order of administration, are: Card I, (*D* 4); Card II, (*D* 2); Card IV, (*D* 1); Card III, (*D* 2); Card V, (*D* 4); Card VIII, (*D* 2); Card VI, (*D* 4); Card IX, (*D* 1); Card VII, (*D* 9); Card X, (*D* 1).

The inkblots were presented to *S* using standard Rorschach instructions except for the following changes: (a) *S* was instructed to look at the whole blot portion in giving his response; (b) the blot was removed as soon as *S* gave one response; (c) if *S* started to rotate the card he was asked to keep it in the position in which it was handed to him. When *S* had given one response to each of the blots, he was told that he would be asked to go through them again, and this time he was to tell *E* what else each blot could be. Following this second series, the blots were again presented to *S* with the same instruction to tell *E* what else each could look like this time. After this third series, an inquiry was conducted in the usual manner, card by card. Thus, three responses were obtained from each blot, making a total of 30 responses given by each *S*.

All Rorschach records were scored according to the

Klopfer scheme. Using ten randomly selected records, interscorer reliability was found to average .85 for all scoring categories. The average agreement for the ten records was 83 per cent. Because of the relative infrequency of the various shading responses of the Klopfer type, all such responses were scored as *Sh*. Thus, seven determinant categories were scored, namely *M*, *FM*, *Fm*, *F*, *Sh*, *FC*, and *CF*. Because *C* responses occurred in only two records, they were treated as *CF* responses for our present purposes.

As mentioned previously, two general measures of complexity were derived from the Rorschach, i.e., determinant complexity and content complexity. Since no one basic measure of either determinant or content complexity has been developed, several measures of each were used. Determinant complexity was measured empirically in four ways by counting: (a) the total number of different determinants used by *S* in his 30 responses; (b) the total number of cards on which *S* repeated the same determinant at least two times; (c) the total number of cards on which *S* repeated the same determinant three times; (d) the total number of cards on which *S* gave three different determinants.

It is apparent that high scores on *a* and *d* and low scores on *b* and *c* reflect greater complexity. Content complexity was measured specifically in two ways by counting: (a) the total number of responses which were repeated by *S* on the same card; (b) the total number of cards on which a response was repeated. Low scores on both the *a* and *b* content measures reflect greater complexity. Interscorer reliability coefficients on both of these content-complexity measures, using ten randomly selected protocols, were .84 and .86 respectively. The distributions of scores of the NDP and Rorschach complexity measures are presented in Table 1.

RESULTS

The two experimental hypotheses were tested by correlating each *S*'s NDP score (complexity of perceptions of people) with his complexity scores on the modified Rorschach. All complexity scores were found to be approximately normally distributed (Table 1).

TABLE 1
DISTRIBUTIONS OF SCORES OF THE NDP AND
RORSCHACH COMPLEXITY MEASURES (*N* = 40)

	Mean	Median	<i>SD</i>	Range
NDP	15.52	16.63	3.41	9-20
Determinant complexity				
a	5.97	5.99	0.76	4-7
b	7.65	7.70	1.60	4-10
c	2.38	2.10	1.91	0-7
d	2.30	2.32	1.55	0-6
Content complexity				
a	7.55	7.00	4.67	0-26*
b	3.25	3.50	2.52	0-9

* Includes one subject with extreme score of 26.

Using the Pearson product-moment coefficient, each of the six Rorschach complexity measures was correlated with the NDP measure. These findings are presented in Table 2.

It will be noted in Table 2 that Hypothesis I is supported by significant correlations between the NDP measure and each of the four Rorschach determinant-complexity measures. Similarly, Hypothesis II is supported by the two significant correlations between the NDP measure and the Rorschach content-complexity measures. It is apparent that the various Rorschach measures of complexity are interrelated because of the way in which they were empirically derived. This would explain the relative homogeneity of the various correlations presented in Table 2. Inspection of Table 2 indicates the Rorschach content measures of complexity and the Rorschach determinant-complexity measures correlate about equally with the NDP measure.

Complexity of response to two stimulus realms, i.e., people and inkblots, has been measured in this study by means of verbal behavior. Consideration of the verbal measures used might suggest that a variable such as verbal intelligence or fluency might account for much of the interrelatedness of behavior on these two perceptual tasks. In order to analyze this possibility, each *S* was given the vocabulary subscale of the Wechsler-Bellevue (Form I). These vocabulary scores were then correlated with the various Rorschach complexity measures and with the NDP measure. It was found that there were indeed significant correlations between four of the Rorschach complexity measures and the vocabulary scores. The absolute range of these relationships was from $r = .37$ to $r = .46$, the most pronounced correlations being found between the content-complexity measures and vocabulary. On the other hand, no correlation ($r = -.01$) was found between the NDP measure and vocabulary scores. Thus, we may assume that verbal facility, to the extent it is measured by vocabulary, is not responsible for the intercorrelations obtained in testing the experimental hypotheses. Indeed, when we partial out the effects of vocabulary on the relationships between the complexity measures, we find the correlations are increased, although not significantly. Parenthetically, the fact that the Rorschach measures correlated higher with

TABLE 2
INTERCORRELATION OF RORSCHACH-DETERMINANT AND
CONTENT-COMPLEXITY SCORES WITH COMPLEXITY
SCORES IN THE PERCEPTION OF PEOPLE

($N = 40$)

	NDP	Direction of Prediction	p^*
Hypothesis I			
a. Number of determinants used	.27	positive	.05
b. Number of cards with two determinants repeated	-.40	negative	.005
c. Number of cards with three determinants repeated	-.39	negative	.01
d. Number of cards with three different determinants	.42	positive	.005
Hypothesis II			
a. Number of responses repeated	-.44	negative	.005
b. Number of cards on which a response was repeated	-.50	negative	.005

* One-tailed test.

the intelligence measure than did the NDP measure might be related to the fact that just as every *S* had to define the same words on the vocabulary test, every *S* perceived the same blot stimuli. On the Rep Test, however, each *S* perceived persons different from those perceived by other *S*s.

Do any of the traditional Rorschach determinant categories bear any relationship to the *S*s' ability to achieve complexity of response in their perceptions of people? In terms of standard Rorschach rationale, it might be expected that the *M* or human movement response would be related to complexity of the perception of people. In the first place, *M* is a perception in which a person is the content. Secondly, a common interpretation given the *M* response is that it involves more internalized, introverted, and possibly imaginative ways of approaching problems concerning the self and other. The results of correlating the various Rorschach determinants with the NDP measure are in line with this view. The correlation between *M* and NDP was .31 ($p = .05$, two-tailed test). Further, this was the only determinant which correlated with the NDP measure at a level significantly different from zero, the other correlations ranging from .04 to $-.06$. Thus, it would appear that the greater the tendency to perceive humans in inkblots the greater is the tendency to invoke more variable or complex perceptions of people on the sorting task.

DISCUSSION

The results of this study suggest that some degree of generality in the complexity of Ss' behavior can be demonstrated using two perceptual tasks involving personal and non-personal stimuli. It is assumed that this generality of behavior is referable to the complexity of the cognitive system of the individual. That is, the manner in which an individual structures and cognizes one realm of events bears some relationship to how he structures another realm of events. It must be emphasized, however, that this is not a chance sampling of events. In the present study, the underlying formulation has been that the individual's learning experiences in the realm of interpersonal relationships provide the basic core from which his cognitive system for construing the world is developed. The fact that there is some generality from the personal to the nonpersonal realms in the cognitive behavior of individuals suggests the trans-situational pervasiveness of some sort of cognitive or perceptual "attitudes" in human behavior, as other workers have posited (10). Our results are also compatible with studies utilizing the Rorschach as a measure of perceptual rigidity (5, 6, 7, 8).

The theoretical problem remains of relating a concept such as the complexity of the cognitive system to the behavior of the individual. Clinically, we are often impressed by the manner in which patients may generalize behavior from one situation to another, perhaps "unrelated," situation. Reacting towards others "as if" they represented some earlier interpersonal relationships has been stressed by Sullivan in his concept of the "parataxic" mode of behavior. What is involved here is the way in which a patient conceptualizes other persons of importance to him, the way in which gradients of similarity and dissimilarity develop in perceiving others, and how subtle or crassly stereotyped the patient may be in these perceptions. These are among some of the behaviors which would seem to be related to such a formulation as the complexity of the individual's cognitive system. A beginning has been made in this direction by relating the cognitive-complexity variable to the predictive accuracy of the individual's behavior (3).

It is important that the methodological problems in obtaining an over-all measure of

the degree of differentiation in one's perceptions of others be recognized. Although it has been demonstrated that an adequate degree of stability exists over time for the measure of cognitive complexity in the perception of people, certain basic problems are posed. For example, the sampling of the people involved in measuring the complexity of perceptions of others is crucial. They may be peers, family members, old, young, same sex, or opposite sex. Many people may be sampled, or only a few. It is apparent that generalization of findings such as those obtained in the present study should be limited by a consideration of these factors.

The use of the modified Rorschach blots in this study underscores a recognized but neglected area of investigation concerning Rorschach behavior, namely, the importance of variability of response both as to determinants and content. The fact that it was possible to demonstrate significant relationships between inkblot-response variability and degree of differentiation in perceiving others suggests the importance of such an approach to Rorschach performance. While perseverative and repetitive behavior on the Rorschach has received most clinical and research attention relative to certain diagnostic patterns, such as organicity and mental deficiency, the results of this study indicate that conceptualizing and analyzing response variability or complexity as a function of the general cognitive system of the *normal* individual may add an important dimension in incorporating Rorschach behavior into a systematic theoretical framework. In addition, the finding that content complexity is *at least* as related to complexity in perceiving people as is determinant complexity, is consistent with the increasing emphasis being placed upon content interpretation in Rorschach practice (13).

SUMMARY

The assumption is made that fundamental to an understanding of human behavior from a perceptual viewpoint is the underlying structure of the individual's cognitive system with which he actively interprets his social world. The structure of this system can be delineated as to its relative complexity-simplicity. It was reasoned that the generality of cognitive complexity, operationally defined by

response variability, could be measured between a perceptual situation involving persons as stimuli and a perceptual task using inkblots as stimuli. Two experimental hypotheses were derived: I, Significant positive relationships exist between the complexity of Ss' perceptions of people and the complexity of their perceptions of inkblots relative to the *determinants* used and II, Significant positive relationships exist between the complexity of Ss' perceptions of people and their perceptions of inkblots relative to the *content* elicited. A measure of cognitive complexity of persons was derived from a concept-formation sorting task originally devised by Kelly (9). The stability of this measure over time was demonstrated. A modified Rorschach inkblot task was used to measure cognitive complexity in the non-personal stimulus realm. The sample consisted of 40 undergraduate college males.

Significant relationships in the predicted direction were found for both experimental hypotheses, using several indices of complexity of response in the Rorschach situation for determinants and for content. Although intelligence correlated significantly with the inkblot measures of complexity, it appeared to play no role in producing the generality of cognitive complexity which was found. Of the various Rorschach determinant categories, only *M* correlated with the measure of complexity of perceptions of others at a level significantly different from zero. The theoretical and clinical significance of a variable such as cognitive complexity is discussed.

REFERENCES

1. BARTLETT, F. C. *Remembering*. Cambridge, England: Cambridge Univer. Press, 1932.
2. BECK, S. J. *Rorschach's test*. Vol. I: *Basic processes*. (2nd ed.) New York: Grune & Stratton, 1949.
3. BIERI, J. Cognitive complexity-simplicity and predictive behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 263-268.
4. BIERI, J., & BLACKER, E. External and internal stimulus factors in Rorschach performance. *J. consult. Psychol.*, 1956, **20**, 1-7.
5. COWEN, E. L., & THOMPSON, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165-176.
6. ERICKSEN, C. W., & EINSTEIN, D. Perceptual rigidity and the Rorschach. *J. Pers.*, 1953, **21**, 386-391.
7. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
8. JOHNSON, L. C., & STERN, J. A. Rigidity on the Rorschach and response to intermittent photic stimulation. *J. consult. Psychol.*, 1955, **19**, 311-317.
9. KELLY, G. A. *The psychology of personal constructs*. Vol. I: *A theory of personality*. New York: Norton, 1955.
10. KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception—an approach to personality*. New York: Ronald, 1951. Pp. 328-355.
11. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.
12. PIAGET, J. *The construction of reality in the child*. New York: Basic Books, 1954.
13. SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing*. New York: Grune & Stratton, 1954.
14. TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189-208.
15. VERNON, M. D. The functions of schemata in perceiving. *Psychol. Rev.*, 1955, **62**, 180-192.

Received October 7, 1955.

THE EFFECT OF STIMULUS AND BACKGROUND FACTORS ON THE VOLUNTEERING RESPONSE

MILTON E. ROSENBAUM¹

University of North Carolina

IN AN earlier experiment (2) the effect of permitting subjects (Ss) to observe the response of other people to a standard invitation to volunteer for a task before the same invitation is tendered to the Ss was studied. It was found that Ss' acceptance of the invitation was significantly greater if they observed another person accepting the invitation than if they observed rejection of the invitation or were not permitted to observe the response of others in a control condition. Significantly less acceptance occurred when rejection by others was observed. These findings were interpreted as indicating that behavior within social situations is the product of a judgmental act in which S is making use of all effective portions of a stimulus complex to determine behavior appropriate for him.

Deriving its orientation from Helson's theory of adaptation level (1), the present experiment elaborates the former by systematically studying the contribution of the stimulus in the immediate focus of attention in conjunction with presentation of background stimuli in the form of the responses of other people.

Basic design. The general plan of the study called for creating test situations within which invitations to participate in a psychological experiment that varied in strength were tendered to Ss after they had seen the reaction of another person to the same request, with the observed reaction also being subjected to systematic variation. Three stimulus requests were used, designed to vary along an intensity continuum in terms of their capacity to produce a volunteering response. Positive and negative backgrounds and a control condition or neutral background were introduced to de-

termine the rate of volunteering in conjunction with each of the stimulus requests.

Hypotheses. The variations of stimulus request strength and the character of social background mentioned above lead to the following hypotheses: (a) willingness to volunteer is positively related to stimulus-request intensity, and (b) willingness to volunteer is directly related to the behavior of others who are observed responding to the same invitation.

METHOD

Setting and subjects. The study was conducted during the regular Fall semester of 1953-54 at The University of Texas. The settings were the reading rooms of two large university libraries. The Ss, numbering 135 males, were occupants of the library going about their everyday school affairs. Both libraries were divided into six geographical areas. On each experimental occasion one of these locations was selected by a counterbalanced plan, and those students were chosen to be Ss who were seated in positions in the designated area that permitted the seating of an experimenter's (E's) assistant next to them. Fifteen test Ss were employed under each of the nine conditions to which they were randomly assigned.

Assistants. Fifteen assistants participated and each served on six different testing occasions, once under each of the six experimental conditions requiring an assistant, to which they were randomly assigned. The assistants were all males recruited from undergraduate and graduate psychology courses. On each testing occasion the assistant was instructed concerning the response he was to give. He was not informed in advance of the intensity of the request that would be used on any specific test occasion.

Experimenters. Two male Es conducted the study proper. Both were graduate psychology students. Their administration of the experimental conditions was randomized over the nine conditions. Although one of them was aware of both the request intensity and the background that would be used in any given test situation, the other E was aware of the request intensity only. He did not know what the assistant's response would be until the response was actually given. A

¹ This report is based upon a portion of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at The University of Texas. The writer is indebted to Dr. Robert R. Blake for his guidance as thesis advisor. Appreciation is also extended to Dr. Harry Helson for theoretical and interpretive assistance. Special gratitude is owed Joseph A. Marin for his splendid cooperation in collecting the data. Dr. John Thibaut and Dr. W. Grant Dahlstrom offered valuable advice in the preparation of this paper.

² It is possible that one of the Es might have induced different rates of acceptance and rejection of the invitation than the other. A comparison of the frequencies of acceptance and rejection obtained by the two yields a χ^2 of .552 which with one degree of freedom indicates a p value greater than .30 and demonstrates that no significant difference occurred in the manner in which the two conducted their roles in the experiment.

continuing observational check was made to insure the consistency of the behavior of the two *Es*.²

Development of intensity differences in the stimulus-requests. A pilot study was conducted in order to devise stimulus-requests that would vary along the dimension of strength or intensity. By using the frequency of acceptance as a measure of the strength of the stimulus-request it was possible to determine the frequency with which a particular form of request would produce the act of volunteering in the context of a neutral field. After a number of preliminary trials with different versions of the stimulus-requests, three that satisfied the requirement of representing discriminably different points on the stimulus continuum were selected for use in the experiment proper.

Each of the requests dealt with a need for participants in a psychological experiment, but they differed in terms of length of time implied to complete the task and the compellingness of the phrasing. The high intensity request was phrased as a plea for subjects to help complete dissertation research. The medium intensity request was a matter of fact statement of a need for *Ss* for departmental research. The low intensity request was stated in a form presenting a relatively weak desire to have *S* participate in an experiment but discouraging him from doing so.

Creation of background differences. Background was created by the reaction of a person invited to participate in the presence of the naive test *S*. By bending over between the assistant and the test *S*, *E* first approached the assistant who had been instructed as to the response he should give. Then the test *S* was approached and requested to participate in the experiment. The general procedure used to create each of the background conditions was as follows:

Positive background. To the invitation to participate in the experiment, the assistant responded by saying "Okay." Instead of acting on the assistant's reaction by leaving the library immediately, *E* asked the assistant to wait standing beside his seat. He then invited the test *S*. This procedure did not allow the test *S* to determine the actual amount of time that would be required to complete the experiment.

Only one test *S* was approached at any given time. If he accepted the invitation after the assistant had done so, *E* escorted them both to an experimental room which was located nearby. The assistant and the test *S* were administered a rating scale which allowed them to indicate the strength of their desire to participate in the experiment. A level-of-aspiration task was administered to the test *S* who upon its completion returned to the library. The level-of-aspiration data are not considered in the present study. If the test *S* rejected the invitation, he was asked to fill out the rating scale at the library table. The assistant and the experimenter then left the library together.

Negative background. The assistant responded to the invitation by saying, "No, I'd rather not," and he turned back to his books. The *E* then presented the invitation to the test *S*. If he accepted, he was escorted from the library and was given the rating scale and the level-of-aspiration task as described under the positive condition. If the test *S* rejected the invitation, he and the assistant were given the rating scale to fill out. The

E then left the library. After a reasonable time the assistant left the library also.

Neutral background. The *S* was approached directly and tendered the invitation. He was not given the opportunity to observe the reaction of another person under this condition. The *E* acted upon *S*'s response appropriately and administered the rating scale.

The experimental situations began on the hour with selection of a position at which the assistant could be seated. At a quarter past the hour, after interaction with the first test *S* had been completed, a second experimental situation was begun. Another location was selected where an assistant could take a seat next to another naive *S*. The manipulation of the background conditions was conducted in a manner consistent with the naturalness of the setting. There were no indications that deception was suspected. The high degree of interest of the volunteering *Ss* in the level-of-aspiration task leads to a belief that the manipulations were seen as a usual process in recruiting *Ss* for psychological experiments.

Measurement of response. Frequency of acceptance or rejection of the invitation serves as one basis for judging the influence of the stimulus and background factors, but the measure of chief interest consists of a rating scale devised for the purpose of measuring the intensity of feeling about volunteering. It was introduced because subjective evaluation of responses in the earlier experiment indicated that *Ss* accepted or rejected the invitation with different degrees of willingness and unwillingness. This cannot be adequately expressed by means of the acceptance-rejection dichotomy.

The scale consisted of thirteen categories, the seventh category existing only theoretically since the conditions did not permit an undecided response. Any single *S* was administered only one half of the scale. If his response was positive, he was given only the portion measuring degree of willingness to volunteer. An *S* whose response was negative was administered the portion dealing with degree of unwillingness to volunteer.

RESULTS AND DISCUSSION

Comparison with the earlier experiment. The medium intensity stimulus was almost identical with the standard invitation used in the earlier experiment with only a minor change in wording in terms of the time implied. Presentation of this stimulus in conjunction with the background conditions essentially constitutes a replication of the earlier experiment. It is of interest then to compare the results to check the reliability of the previous findings. Significant differences were found in the past among the background conditions when frequencies of acceptance and rejection were evaluated by means of χ^2 . Greatest acceptance occurred under the positive background and least under the negative background. The data necessary

TABLE 1
FREQUENCIES OF ACCEPTANCE AND REJECTION UNDER
THE MEDIUM INTENSITY STIMULUS-REQUEST

Background Condition	Acceptance	Rejection
Positive	12	3
Neutral	7	8
Negative	0	15

TABLE 2
RESULTS OF THE ANALYSIS OF VARIANCE OF THE
RATING SCALE DATA

Source of Variation	df	Mean Square	F
Background	2	57.63	10.16*
Stimulus	2	118.69	20.93*
Interaction	4	9.73	1.72
Within	126	5.67	

* Significant at the .01 level.

TABLE 3
MEAN SCALE VALUES AND FREQUENCIES OF ACCEPTANCE AND REJECTION BY STIMULUS-REQUEST INTENSITY AND BACKGROUND CONDITION

Background Condition	Stimulus-Request Intensity								
	High			Medium			Low		
	Mean Scale Value	Frequency		Mean Scale Value	Frequency		Mean Scale Value	Frequency	
		Accept	Reject		Accept	Reject		Accept	Reject
Positive	9.40	12	3	9.07	12	3	7.53	6	9
Neutral	9.13	12	3	7.40	7	8	4.60	0	15
Negative	8.60	11	4	5.47	1	14	5.33	1	14

to evaluate the similar aspects of the present experiment are presented in Table 1.

For the differences among the three background conditions, the χ^2 is 19.86 which with two degrees of freedom is significant beyond the .001 level of confidence. It is apparent therefore that the replication within the present experiment of the conditions of the earlier experiment yields the same result indicating that response to a medium intensity stimulus-request is directly related to the character of the background condition.

The relationship of stimulus and background to volunteering. The experimental procedure permitted an analysis of variance in a 3×3 factorial design. For this purpose the rating-scale data were used as the measure of response. Bartlett's test yielded a χ^2 of 10.72, $p > .05$,

supporting the assumption of homogeneity necessary for the analysis of variance.

The analysis of variance is summarized in Table 2. the within-sum-of-squares serves as the appropriate error term. Both the stimulus and background variables are seen to be significant in their effect on the response elicited in the predicted direction: the higher the intensity of the stimulus-request, the greater the willingness of the test *S* to volunteer. On the other hand, the more conducive the background conditions toward the act of volunteering in terms of the observed behavior of another person, the greater the willingness of the test *S* to volunteer. No significant interaction is indicated between the two variables. Both hypotheses therefore are supported by the results.³

The frequency of volunteering apart from degree of willingness is of interest from a prac-

tical standpoint. For comparative purposes Table 3 presents both the mean values obtained from the rating-scale data and the frequencies of acceptance and rejection of the invitation that appeared under the various conditions. χ^2 analysis of the frequency data supports conclusions similar to those obtained from the analysis of variance of the rating-scale data. χ^2 values significant beyond the .01 level are obtained for the relationship of both stimulus-request intensity and background

³ Further analysis in order to evaluate the differences that appear between the levels of each variable was conducted by means of the *t* test. It is found that for the most part the differences are significant at the .01 level. The single exception occurs in the comparison of the neutral background with the negative background for which the *p* value is greater than .10.

condition to volunteering. Again there is no interaction effect. The rating-scale measure is more sensitive than actual volunteering, however, to the effect of the variables. Under three sets of combined stimulus and background conditions in one case, and in two sets in another, the frequencies of acceptance of the invitations are identical, while the rating-scale data clearly differentiate the conditions.

Although the present results clearly indicate that knowledge and control of immediate stimuli and background stimuli lead to predictability of a simple social response with a high degree of confidence, this experiment did not manipulate or control a third major class of contributing variables: "residual stimuli" (in Helson's terms) deriving from such factors as genetic predispositions, prior experiences, attitudes, abilities, etc. The contribution of this source to the unaccounted-for variance in the experimental results was strikingly demonstrated in the response of one S. This S, under the condition least conducive to volunteering, the low stimulus-request intensity and negative background condition, accepted the invitation to volunteer, and on the rating scale checked Category 13, "would not miss for anything." He explained spontaneously that he was a freshman planning to major in psychology and that this was the first opportunity he had been offered to be in direct contact with any type of psychological work. Undoubtedly

responses given by other Ss were to some extent conditioned by internal factors such as an examination the next hour, fear of psychological experimentation, lack of interest in their library studies, and many other individual characteristics. Major problems exist in the identification and control of the major sources of stimulation leading to the appearance of social responses but the present results suggest the utility of the threefold classification discussed above.

SUMMARY

The response of Ss made up of university library readers to an invitation to participate in a psychological experiment was studied under varying conditions of stimulus-request intensity and background stimulation presented in the form of responses of other people in the same situation. The background stimulation was controlled by use of assistants to the experimenter. Both stimulus and background factors were found to be significantly related to the volunteering response.

REFERENCES

1. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
2. ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, 50, 193-196.

Received September 20, 1955.

CASE REPORTS

RUMOR IN A PRIMITIVE SOCIETY

RAYMOND FIRTH

London School of Economics and Political Science

IN a small-scale "primitive" society of the type usually studied by anthropologists, what is the currency of rumor, and how do its processes compare with those of rumor in more sophisticated societies? These questions have not yet been studied. Rumor is a subject of which most anthropologists in the field seem to have had much experience, generally and as affecting themselves. But, curiously, they have written little about its formation, its incidence, and its social repercussions. This essay is meant to open up the problem by analysis of data from the Polynesian island of Tikopia in the Western Pacific.¹ It is also hoped that the material will be of interest to psychologists.

What are the main characteristics of rumor? They are these: talk or report of hearsay kind, not original expression; general currency or spread of such report through a social group; assertions of doubtful accuracy or unverified.² For an anthropologist many questions are involved in rumor. What is its material or content in a primitive society? What themes are most commonly involved? How far are they of individual, how far of social concern? Are these themes few and repetitive, or do they show a wide range of variation? Is the treatment of them relatively prosaic, correspond-

ing fairly closely to ordinary verified experience, or is it experimental, with much fantasy? Other questions arise regarding the originator of rumor, the agents who spread it, and the efforts made at its verification or disproof. There is also the time factor—are rumors ephemeral or persistent; are they slow or do they come in rapid succession?

Finally, what is the functional significance, if any, of rumor in relation to the structure and organization of the society where it occurs? Its negative, disruptive functions often seem fairly clear; the very use of the term is ordinarily pejorative. But can any more positive social functions be discerned in its operations? Does rumor ever promote group solidarity without apparent scapegoat-creation or other social damage? Does it provide individual defense mechanisms against group pressures? Can it be a means of group mobilization for social action? In the following brief account I try to provide interim answers to some of these questions.

As a general proposition, I would hold that rumor has not just purely negative social functions, and that it is not simply the product of idle curiosity or fantasy, but serves as a social instrument, helping groups or individuals to gain their ends.

Reporting, Invention, and Rumor

When I was in Tikopia in 1928-29 and in 1952, though I did not make any detailed study of rumor, on both occasions I noted many of the rumors encountered and their relation to social affairs. My assistant in 1952-53 did this also. Our records are the basis of the following analysis.

First consider the question of accuracy. The line between *news*, ideally the reporting of verified events, and *rumor*, the reporting of unverified events, is a very difficult one to draw, as every newspaper recognizes. At some point between eyewitness and recipient of news in any society there must be some degree of trust in the veracity and the accuracy of the informant. This is one of the requirements of

¹ I am indebted greatly to the Behavioral Sciences Division of the Ford Foundation for the help of Mr. J. Spillius, my research assistant, in contributing toward this material, and for secretarial services in preparation of the article. General data about Tikopia are given in Firth (2, 3).

² I drafted this essay originally without seeing the interesting analysis by Allport and Postman (1). But the above criteria, which seem to represent the lowest common denominator of the concept in popular usage, fit fairly closely Allport and Postman's more precise definition: "rumor is a specific (or topical) proposition for belief, passed along from person to person, usually by word of mouth, without secure standards of evidence being present" (1, p. 9). I have been encouraged in the publication of this essay by discussion with Gordon W. Allport, Jerome S. Bruner, David M. Schneider and other colleagues at a meeting of the Anthropology Colloquium and Social Relations Colloquium at Harvard, where the material was first presented.

social living. If there should be no trust, then nothing but eyewitnessing of events could serve as a basis for action. If there should be trust but no veracity or accuracy, then social living would soon break down through lack of correspondence between account and subsequent experience. In any society, however, there can be only an approximation to the ideal. Facts offered to the eye may be incomplete, the observer may see only part of the event which he wishes to describe. And particularly if any attribution of *intention* be included in the report, the interest of the observer himself may lead him to misrepresent it. With the recipient then, there always tends to be some amount of reserve as to the correctness of report. All this is a statement of sociological first principles. How does it apply to the case of Tikopia?

In Tikopia, a vast amount of ordinary news is passed on by word of mouth daily. This is so especially within any village, and between villages in the same district. Folk are continually traveling up and down the beach and along the inland paths, and exchange of news is part of regular social intercourse, as it tends to be in any rural area. Between districts there is less contact. Formal visiting is not very frequent between their inhabitants. When it occurs, as for a dance festival or some other large-scale ceremony, or when a vessel arrives from overseas, the visitors are apt to remain separate from the local residents, to hang about together under the trees or to sit together in a house. But even then there is plenty of exchange of information between the parties. And in everyday life people of different villages throughout the island meet a great deal, as in the cultivations, which are dispersed, not concentrated in district ownership. The great amount of accurate reporting arising from these social contacts must be set as a background to the prevalence of rumor.

The Tikopia have an avid interest in all such news. They are eager to hear about events in the outside world and also about events, including movements of people, in their own community. An arrival from one side of the island at a ceremony on the other is usually asked first for news. When in 1928-29 I came regularly over a gap in the hillcrest from my house in Faea to the temple of the chief of Kafika in Uta I used to be asked every day

by him at once "Any news from Faea?"—and I had to tell him about the fishing, any illness or death, what the local chief there was doing, and so on. In July, 1952, when the same old chief was spending the night at his sacred glade in Somosomo inland, I joined him there early next morning. "Is there not any news from the strand?" he inquired immediately. "No" was the reply, to his evident disappointment. The term which I have translated as "news," *taranga*, is also that generically for "speech," illustrating the verbal nature of communication in Tikopia.³ This interest in news is presumably a correlate of the small size and isolation of the community. The color and richness of Tikopia personal experience are derived largely from consideration of the minutiae of social existence in their tiny area. Yet this intense curiosity is matched to a surprising degree by an effort at detachment. The news is received as conveyed by individuals, but often treated as if it stemmed from a generic social source.

As may be expected from a linguistic usage which identifies news with speech, the Tikopia make no clear-cut categorical distinction between news of verified, accurate type and rumor, unverified and often inaccurate. But indication of some differences is often given by the expression *a taranga o faoa*, meaning "speech of the crowd" or "speech of people in general," or by the introductory phrase "*E ati . . .*," "It is held that. . . ." The very fact that no precise attribution of origin is given implies an element of vagueness and the possibility of inaccuracy. Such an expression conveys a derogatory note, a judgment or evaluation in advance of investigation—or at least a suspension of acceptance of the story as correct. It also dissociates the speaker from the statement in advance. An analogous expression is "*Faoa e muna . . .*": "people say. . . ." These are the nearest equivalents to our "Rumor has it. . . ."

Despite the lack of a very specific Tikopia expression for rumor as distinct from news, there is a very great interest in the truth of accounts received. The question is often asked, is the story true (*maori*) or false (*loi*)? Is it correct (*tonu*) or not? Such an attitude may

³ Only recently and still very rarely have Tikopia communicated by letter or other indirect means, such as radio.

be also expressed when a story is passed on. The narrator in telling it may add "We don't know whether it is true or false," or "We don't know whether it is correct or not." This indicates not only suspension of judgment on the veracity or on the accuracy of the account; it also indicates that the narrator himself is not committed, and cannot be held responsible for the story afterwards. Moreover, action may be suspended until verification is received. An example of this occurred in October, 1928. It was reported early in the morning that Pa Maevetau, a man of rank in Rofaea, had died. I was living in Matautu about half a mile away at the time and noted that all the people there waited to hear if the report was true (*maori*) or not. Then confirmation came by ceremonial whooping and firing of guns from Rofaea—a token of some important event. Immediately there was great excitement in Matautu. People went about whooping; they dropped all work; and prepared to form mourning parties. An immediate decision was taken to postpone a dance which had been arranged.

Some further features of the Tikopia speech background are relevant to the discussion of rumor there: the people's interest in storytelling; the accuracy of their memory; their capacity for recognized fantasy and invention.

The Tikopia are great narrators. They delight not only in hearing news as items of information; they also delight in giving and hearing a presentation of news in elaborate aesthetic form, with dramatic emphasis. They dwell on incidents, the narrator taking time to explain in particular his own emotions and thoughts as an event takes place. Such narratives are normally presented as a record of actual events. In this the reproduction of remembered items may be often extremely accurate. In 1952 I heard a Tikopia describe to others many details of my movements in 1928-29. A stone on which I had sat under a tree in a pause in a walk twenty-three years before was pointed out. A funeral gathering was told by one man of a song which I had not finished writing down at his dictation then (the remaining stanza was dictated to me on the spot) and so on. And in 1928 I was given details of the visit of H.M.S. "Mohawk" thirty years before, including loss of one of her anchors off Tikopia—a story which on

checking recently (in 1955) I found to be quite correct. The Tikopia are fully capable of accurate accounts from memory.

As regards fantasy, there are a number of Tikopia traditional stories of "folk-tale" character, some brief, many long and elaborate. These are recited informally and often fragmentarily, on no set occasion. One may be told when a group of people are resting after a ceremony, when members of a household are waiting to go to sleep at night, or when a grandparent is looking after a grandchild. Such tales, though their truth may be a matter of suspended judgment, or they may even be regarded as fiction, are usually not regarded as specifically *invented*. Only once did I hear a man tell a story which, with some amusement, he declared to be an invention of his own.

Where the element of personal invention does come out in a very marked fashion, though not so acknowledged by the Tikopia themselves, is in the dissociated state of spirit mediumship. In this condition a Tikopia medium, especially if in a light state of dissociation, may give a long narrative dealing with the adventures of a spirit. But this is regarded by Tikopia as falling in the category of a true recital, granted that spirit mediums as individuals are recognized as being capable of untruth or unverified and inaccurate pronouncements.

On the whole, it may be said that the Tikopia recognize no great field for free fantasy—with one exception, that of song. Here a person is allowed free rein to his inventions and the way in which human relations are handled in songs has often considerable subtlety. There is also a wide range of analogy and symbolism in the songs. (2, pp. 196-197, 289-302, 520-523). But they are apt to be stereotyped and they lack the soaring of the imagination which characterizes songs in some other primitive communities. In perceptiveness about human relations the Tikopia are highly skilled, but they are not very fertile in imagination in fields outside their immediate experience. Even within such experience they do not seem to indulge in much deliberate fiction or in much speculation about possible differences from ordinary behavior. On one such occasion one of my friends observed in a classical style, "If mosquitoes were as big as the birds which fly above, when they fed on men, we

should disappear." But this is the only remark of the kind which I recorded. I think it would be fair to say that free fantasy of an *intellectual order* is not common among the Tikopia. What is common is free fantasy of an *emotional order*, and this forms the basis of much rumor.

General Type of Experience Used by Tikopia Rumor

An important question in considering rumor is what type of experience does it use? Allport and Postman (1) have pointed out how in the formation of rumor psychological processes of leveling of detail by omission, and conversely, of sharpening of detail by selective retention, are very relevant. Assimilation (by cultural and individual selection of perception, memory, and reproduction) also converts an incident into narrative material for more general transmittal. Such processes would seem to be valid as a description of what happens in Tikopia.

But since a great deal of such process is assisted or even short circuited by common acceptance of *signs*, it is pertinent to ask what the Tikopia do. In our Western culture some signs are selected from natural phenomena, others from cultural phenomena. Black clouds are a sign of rain; the whistle of a railway engine in open country is a sign that a train is approaching a level road crossing. Both natural and cultural signs may have social implications. Rain affects human crops, and an engine whistle affects human traffic across the railway lines. But with us, ordinarily a "natural" sign is an indicator of natural occurrence, and a sociocultural sign, an indicator of social occurrence. So is it also with the Tikopia. Among natural indicators, black clouds likewise are held to denote rain, other types of clouds denote wind, and so on. As cultural indicators, smoke rising from trees gives a sign that an oven is being made ready for a ceremony; or gunfire announces the death of a man of rank. The Tikopia have great interest in such signs or indicators. (They have a term *fakamailonga*, as a generic term for indicator.) But a point of interest in Tikopia perception and reproduction is that some natural events are interpreted as signs of social occurrences. A certain kind of ribbed cloud structure denotes to the Tikopia not merely a kind of weather but also the ap-

proach of a vessel which is as yet invisible below the horizon. Similarly a rainbow or the sound of thunder can be a vessel sign. Even a sneeze can be so interpreted.

An important point about these natural signs or indicators is that while from our Western point of view a social interpretation of them is unrealistic or nonempirical, from the Tikopia point of view this has a fair degree of empirical validity. Hence stories based upon them are relevant to the classification of rumor, but marginal in that classification. A story of a ship coming, based on the sight of a rainbow, is to be treated as a positive error, or mistaken attribution, in which the elements of interest-guided perception lie farther back, as it were, and have become culturally pre-selected and standardized. Moreover, in general such indicators as rainbows, cloud formations, sneezes, are taken not as a basis for the spread of stories but for speculation. They indicate probabilities, but of no high degree. People do not pay them much attention, but wait to see if the prediction is justified. Often the person who makes the identification himself regards the sign with some skepticism, or refers to it in terms of possibility—"Is it perhaps that . . . ?"

It may be noted that despite the keen interest the Tikopia show in the arrival of vessels, and their tendency to seize on indicators of approach, they have not utilized dreams for this purpose. Yet they use dreams fairly freely as indicators for foretelling the sex of a child, or a catch of fish. This may be related to the fact that whereas rainbows, clouds, and even sneezes are patent to outside observation and are checkable, the occurrence of a dream relies solely on individual statement. Hence the signs which it gives are apt to be construed as of individual rather than of public significance—that the person concerned will have the experience predicted. That occasionally a dream may provide the basis for rumor is shown by a case where after a dream a man spread the story that a shark caught by a fishing fleet had been caught by his clan; it turned out that he was wrong. But ordinarily dreams are signs of personal, not public, occurrences.

But if neither cultural signs nor dream experiences provide the stuff of rumor, it is different with the material provided by dis-

sociation. Statements issued by spirit mediums are heard publicly and are regarded not as individual pronouncements of the human personality, but as information or instruction from the spirit world. Hence they are deemed worthy of serious belief. They are the basis of many rumors. In 1928, as the result of the statement by a spirit medium, it was widely believed in Tikopia that if I attended the pagan religious rites I would take away their sacred power of bringing rain and fertility to the crops. Luckily, when I first attended, I came in a heavy shower of rain, no other deleterious consequences occurred, and the rumor soon died down. Apart from such types of material, rumor in Tikopia is based for the most part on misinterpretation of ordinary physical events of a cultural order. Examples of this are given later.

A word may be said here about the agents of rumor in Tikopia. Children there are widely blamed for creating and spreading rumors. In this as in many other primitive communities, they have a very positive role. Often going around in gangs, they form a ubiquitous and semiautonomous social element of limited experience and without much responsibility. There is no doubt that they do in fact generate, spread, and elaborate many rumors, particularly those concerned with approach of vessels and movements of people. But to some extent they are a conventional scapegoat, to whom authorship of rumor is often shifted when its falsity becomes manifest. Many rumors are the work of adults. Support to this, apart from absence of children as narrators, is given by the fact that such rumors are often clearly in an adult sphere of interest. Women seem to be authors of rumors as well as men.

Content of Tikopia Rumor

Externally, the content of rumor in Tikopia is concerned with two basic subjects: the movement of vessels in relation to Tikopia, especially their arrival; and the fate of Tikopia abroad. Internally, it covers a wide range, from the timing of ceremonies and doings of men of rank (including anthropologists) to quarrels, and reasons for illness. In all, from my own records in 1928-29 and 1952, I have extracted 38 specific rumors, and from the records of my assistant come another 36, making a total of over 70 rumors noted over a period of about

28 months. This is a minimal figure, since we did not set out to record rumors on all small details of behavior, or other relatively ephemeral matters. (Practically all these rumors were false.)

The lack of data on minor rumors, relating to ephemeral issues, means that our material does not represent a complete cross section of Tikopia rumor. It is not completely representative for another reason—it is biased in the direction of rumors affecting our own movements and intentions. This is natural—for operational reasons it was important to know what was being said about us or matters in which we were concerned, so as to study how to handle the situation. Such rumors usually came rapidly to us—we were speedily asked, as a rule, if they were true. In addition, our doings in themselves were the object of the greatest curiosity and speculation to the Tikopia, since the anthropologists have been the only Europeans for over a century living on Tikopia for any long period. But I think that by one channel or another, most of the major rumors on Tikopia during our stay reached us—that is, most of those which not only tended to alter the social behavior of the recipients significantly, but also affected the organization of social affairs on a considerable scale.

Examples of minor and of major rumors will illustrate some of the procedures in rumor formation in Tikopia.

Minor rumors. One evening I was told that my assistant was not returning to our house in Ravenga that night, but was going to sleep in the dwelling of his friend, the chief of Tafua in Faea, where he had been during the day. The message was brought to me and others in the house of the Taumako chief in the village where we lived. It came through a small girl, who had been listening to the gossip around the bearer of a note from my assistant to me; the messenger had come to my house, and was waiting for me under a tree when I got back. When I read the note, it turned out to be merely an announcement that my assistant was going to be back late for dinner. This is typical of much Tikopia rumor. The delay in the arrival of my assistant; the arrival of a messenger from Faea—these are put together with a false inference that the intention is stopping the night. The inference is not so stupid or far-

fetches as a Western observer might think. A Tikopia on a visit to the other side of the island would not normally send a special message back to say that he would be *late*. He might send one of his companions, a child or young person, to return ahead of him, with instructions to say that he was following. But a special messenger would be sent almost certainly only to tell his family that he was staying the night away. So Tikopia custom gives some plausibility to the rumor, and the force of the interpretation was added to by the sight of a written document, held by them to be a weighty thing. But in addition there is the element of interest in novelty. There is more fun in speculation about someone's staying the night away than in his returning home. A rumor of similar exaggerative type came to me when a man fell from a breadfruit tree. Some boys told me that evening that he had broken his leg above the ankle. But his nephew, who went and saw him, told me later that he had hurt his back and grazed his shin, no more. The comment in my notebook at the time (November, 1928) was that "rumor is quick to exaggerate in Tikopia."

One further subject of minor rumor which may be mentioned is personal illness. Rumors in this field range widely from assignment of cause to attribution of credit for treatment and cure. One instance of this was a rumor that the illness of my assistant on a certain occasion was cured by a spirit medium—who, in fact, worked no cure since he came to our house for quite another purpose. As regards causes, rumors frequently involve flights into the realms of spirit action, a subject too elaborate for discussion here.

Major rumors; movement of vessels. A major and continued subject of rumor was the movements and arrival of European vessels. The isolation of the island makes the coming of a vessel—in normal times, only once or twice a year—an event of great importance. For a vessel to call is a time of excitement, news, valuable exchange, gifts, and perhaps arrival and departure of persons. It is much desired, for economic as well as social reasons. Hence there is much inducement to false identification—a tiny cloud on the horizon looking like a smudge of smoke, a distant unidentified object, and a cry of "The ship! The ship!" rings out. Such

is the novelty-interest and wish-fulfillment type of rumor.

During the second expedition to Tikopia there were at least a score of rumors of vessels recorded. This was partly a function of the strain and emotional intensity of the period—owing to hurricane and drought a food shortage had developed into a famine, and prospects of food relief were a vital question. It was also related to the presence of our radio telephone set, with fairly frequent messages.

Here are some typical cases. On the 5th of June, 1952, a rumor began to spread in our village about 9 A.M. "A ship has come." Large numbers of people, men and girls, went off to the other side of the island to see and to trade. A ship was indeed expected that day with food supplies, so it was very likely that some person might deceive himself into believing that he had seen its smoke on the horizon. But what was interesting was the lack of skeptical verification. The Tikopia have been deceived innumerable times by such rumors. And each time afterwards they talk about "*Te loi o faca*"—the falsehood of the crowd. But they seem just as gullible next time. On this occasion we questioned our retainers as to the accuracy of the report. One of them said that a villager, one of our neighbors, had gone along and verified the story; it was quite true. I asked if the vessel was still far off. He replied, "Oh, she is standing inshore." But about 10.30 A.M. the news came back that the story was entirely untrue. The disgruntled crowd streamed back.

Four days later the vessel arrived with food supplies. Ten days later again, when many of the supplies had been consumed and there was some wistful hoping that further food might arrive, another rumor began. It was said that a Government vessel would come that day or the next; that my assistant and I had announced this after hearing it on the radio telephone. People should sit in their houses, we had stated, and not go the woods to get food; simply sit and wait for it to be brought by the vessel. Popular support was given to this story, by the fact that for the last two days the peculiar cloud formations supposed to be indicative of approaching vessels had been observed. We ourselves had had our attention drawn to them—for instance by the old chief of Kafika. The basis for this rumor about the ship apparently was that in a news item from the Solo-

mons on the radio the day before, it had actually been mentioned that a Government vessel was proceeding from Honiara, the capital of the Protectorate, to Gizo, a port in the West. This is in the opposite direction to Tikopia, and more than six hundred miles away, but someone in the crowd which always assembled whenever the radio was in action (5, p. 810), had evidently caught a part of the message and garbled it. A day later the rumor took a further form—of criticism of my assistant and me, and of the radio, for (it was alleged) giving wrong information! It was said that we ourselves had *started* the story of the coming of the Government vessel to Tikopia and then it turned out not to be true!

Movement of persons. Another set of rumors was concerned with our alleged movements. One was that I had tried, a year or so earlier (the time varied in different accounts), to return to Tikopia. (This was in accordance with a statement I was alleged to have made when I left in 1929.) But, it was said, I had been prevented by the Government or by some other agency after having got part of the way, and turned back. The bases for this rumor were dual. On the one hand I did, in 1929, express a *wish* to return to Tikopia, though I had said I was uncertain if I ever could. On the other hand, I had actually paid a visit in 1951 to New Guinea. I had no intention on that occasion of attempting to reach Tikopia. But the airline to the Solomons was by way of New Guinea. When in the end after my arrival I told some Tikopia about this earlier trip it provided the basis for the story of a thwarted start for the island. Again, in May, 1952, the story became current that, in one case I, and in another my research assistant, was going to Vanikoro, another island about 120 miles away, in a Government vessel to buy tobacco and return with it to Tikopia. The basis for this was an acute shortage of tobacco in the community at this time; several people had half jokingly suggested to us that we should send out an order for more or go and get more—without regard, of course, to the finance involved. Such rumors are of the wish-fulfillment type. Their genesis lies basically in an attempt to give concrete form to a strong desire. But others, also related to the movements of persons, are essentially an anxiety expression.

The vessel on which the first shipment of

relief food supplies was brought also took away a number of Tikopia who had recruited as plantation laborers. About seven weeks after they had gone the rumor spread that disaster had overtaken this labor force—though no details were given even as to whether it was on sea or on land. Further, it was alleged that my assistant and I had received this news on the radio-telephone and had concealed it. (Hardly necessary to say this was complete fiction.) As a variant of this was the rumor that two men of this force, cited by name, had died. About a week later the story had assumed a different form, that two men had died—one of those cited earlier, but the other of a different labor shipment, at Vanikoro. Again, it was said the report had come by radio, but that we had not made it known.

On the one hand, these rumors were reflections of the fear of Tikopia for their kinsmen away at work, which past experience had shown usually resulted in several deaths (2, p. 42). On the other hand, they were also expressions of the resentment of some Tikopia at our restricted use of the radio-telephone as a news instrument. Here was a magnificent source of news, we were the only channel through which it could come, and instead of furnishing them with frequent information about their kin or any other matter in which they were interested, we just listened to weather reports and more general news which they did not understand, and selfishly engaged in business conversations of our own! Such was the attitude, not very clearly formulated. Indeed, we guessed at this view, rather than actually heard it expressed.⁴ Such rumors, fairly directly related to fears and hopes for kin, are a reflection in part of the Tikopia family structure and its system of sentiments.

Another rumor with personal referent was concerned with the recruitment of the laborers.

⁴ We did make what inquiries we could about the laborers—including report of several deaths—and transmitted news about them, apart from keeping the Tikopia informed of matters of public interest. In particular also, we kept in touch with the Government in Honiara about the approaching food famine. This, even apart from matters concerned with the organization of the expedition, was difficult owing to serious weather distortions and breakdowns, which only the devotion of my assistant to the instrument was able to remedy. Most Tikopia seemed genuinely appreciative of this service, and made this known to us.

At the time when the vessel called I was asked by a man of rank why my assistant was barring men of the district where we lived from being "written down," i.e., having their names entered in the recruiter's book. I asked what he meant. He said that my assistant was "writing down" names of men of the other district to be recruited, but not those of our district. Later, two other men of our village said much the same to me. The genesis of this rumor was an ordinary field inquiry to ascertain the motivations and attitudes of those men who were offering themselves for recruitment, or who had stated they wished to go. Since the vessel had anchored off the coast of the other district, and its boatswain, a most active recruiter, was a Tikopia man from that district, he tended in the first place to recruit or approach men who were his own neighbors or kin. But the traditional suspicion between the people of the two districts coupled with the very limited knowledge of English possessed by most Tikopia, and their ignorance of recruiting procedures and how to get themselves on the list, gave rise to the view that favoritism was being exercised by my assistant.

Rumors based on other fears may have a broader sociological referent. One of these involving alleged movement of both vessel and person was associated with the receipt of relief food supplies—which came mainly in the form of rice. A rumor in July, 1952, after the first relief vessel had been gone some time, was to the effect that one of the most respected and best known of the European priests of the Melanesian Mission stationed elsewhere in the Solomons, had objected to this rice being sent, saying, "Shall it be food for pagan religious rites?" He said (it was alleged) that if another food vessel was sent he would see that it "sank at sea." He had been on Tikopia nearly half a century before, and was known by a number of Tikopia men. One of my old Tikopia friends who had seen this man as a boy asked me "Is he someone the spirits desire? He has 'power'."

In this rumor there emerge several themes characteristic of Tikopia religious thinking. The first is the opposition between Christian and pagan. In this case it is projected on to the distant priest, in actuality a saintly man, wholly innocent of such thoughts as the sinking of relief ships. The second is the translation of that opposition into terms of the outstanding

economic preoccupation of the Tikopia—food, and its use. The close association in Tikopia thought and action between religion and food is seen in the whole pagan cycle of the "Work of the Gods," as also in the Christian argument that the pagan religion is wasteful of food by its offerings and ceremonial accumulations and feasts. The third theme is that of the possession of supernatural power by men of rank. The European missionary's suggested "sinking" of the vessel alluded to was to have been done by supernormal techniques, and not by physical sabotage. Hence, the question of my friend as to whether the missionary was "desired" by the spirits, meant, did he have their support?

Violence and social strife. Traditional structural elements were reflected in another type of rumor concerning the treatment of thieves. The rumor became current, as the food situation worsened, that if the rice shipment did not arrive the chiefs would have given orders that thieves would be killed. I told one of the chiefs of this. He laughed and said "Oh no!" He explained that what had been said was that anyone who thieved consistently should be carried off to work in the Solomons. This, in fact, was the policy of the chiefs, and they did succeed in having several thieves recruited as laborers. But one traditional punishment in such cases was banishment to sea, which was virtually equivalent to execution (4). The rumor here was doing two things: it was reviving this extreme punishment in anticipation; on the other hand, it was assimilating it to the general type of physical rough handling to which thieves, in fact, were traditionally subjected.

A little later we had a repercussion of this rumor in the form of a message from the Government of the Solomon Islands Protectorate, stating that it had been reported on Tikopia authority that if it had not been for the arrival of relief supplies the chiefs would have taken the matter into their own hands and killed thieves. We were asked for comment. We replied that the story was without foundation. But almost at once two other radio messages came from Tikopia police constables at District headquarters addressed to their kinsfolk on the island. Each was from a son to his father. One was: "I heard news stop People there thieving stop advising you and the family not to follow them." The other was: "I heard

news about there stop look after people do not do any harm to them." These messages were interesting from several points of view. They used the father's personal name in address, thereby breaking one of the stringent Tikopia taboos. They were in English through an intermediary. They were certainly the first time that Tikopia had ever used radio as a means of communication with one another. But the most significant point in this context is that both these young men obviously believed in the possibility that the rumor was true. This was partly a reflection of their belief, later shown to be well founded, that violence was not far from the surface of Tikopia public affairs in such a time of severe strain and crisis. It was also partly an indication of the extent of their credence in what was, at that time, a far-fetched suggestion.

As the famine crisis developed rumors of violence became more frequent. They took such forms as this: "An aristocrat was going berserk and was going to take his club and kill commoners." "Two prominent executive officials were going to kill thieves." "Executive officials were about to fight among themselves in a struggle for power." Each of these had become, indeed, a none-too-remote possibility, and each followed the traditional Tikopia cultural pattern.

The rumor about the chiefs having thieves killed involves the conception of the chiefs as a primary source of public order in Tikopia. But it implies more—opposition of interests, as well as the coincidence of interests between chiefs and their families, on the one hand, and commoners on the other. This was exemplified in another rumor—that the chiefs were going to band together and drive out the commoners to seize their lands. This is a theme of long standing in Tikopia. I was asked about it rather anxiously by leaders of the commoners in 1952, but it was then only a fear and demand for reassurance, not a concrete rumor. Later, when my assistant was alone in the island and the food situation had degenerated still more, it became rumor. At his instance there was a meeting of chiefs and executive officials in the Ravenga district where he lived, to discuss plans for the distribution of a further food supply being sent as relief. This meeting was interpreted by a spirit medium of the other district, Faea, as a meeting to decide on the

long-talked-of expulsion of the commoners of Faea, to make them set off in their canoes, leaving their homes forever. As a result the men of Faea bound their heads for war, took down clubs, and conferred excitedly, determined to resist. On their return, the executive officials of Faea were met by armed parties. Seeing how the situation stood, they circulated among the villages explaining what had really occurred. But to allay the fears of the commoners my assistant was asked to go over to Faea in person, and make a normal appearance as a demonstration that nothing untoward was happening. On his way over he was asked on two or three occasions, half jokingly, by women working in the cultivations, if it was true that the aristocratic families were about to expel them from Tikopia. Here the dormant solution of traditional type was evoked once again in time of real crisis.

Recapitulation and Functional Implications

I have shown that with the Tikopia, the content of rumor is closely related to their experience, and that while the total range is wide, the main themes are relatively few and repetitive. While the mode of communication itself is almost wholly verbal, Tikopia rumor has shown itself capable of seizing upon a new medium, the radio, and of incorporating it by fictitious citation into the rumor content. Psychologically, the processes of formation would appear to be similar to those current in more sophisticated societies. But sociologically, the types of experience used vary considerably from those used in the West, e.g., in the public reliance placed on the pronouncements of spirit mediums.⁵ It is difficult in Tikopia, as elsewhere, to identify the agents of rumors. But occasionally an original utterer can be identified, usually, it would seem, someone with a power-axe to grind. The Tikopia practice of tending to avoid or disclaim public responsibility, however, reduces the possibility of this. As regards spread of rumor, its currency in time, and the credence given to it, there is wide variation in Tikopia. But most rumors are ephemeral. They tend to succeed one another in fairly rapid succession, especially when the issue is one of immediate emotional signifi-

⁵ A Western analogy is the statements of soothsayers, astrologers, etc.; but reliance upon them for public action is marginal.

cance—as a famine, or the fate of absent kin. When they persist or reappear, it is because they correspond to some deep-seated structural cleavage. But for the most part they diffuse quickly through the whole community. Whether they do so depends in part on the possibility of check (willingness to check being in itself an index of some skepticism) and in part on whether the rumor is of major or minor import.

I now draw some further implications. To begin with, one can classify Tikopia rumors into two main divisions: those which are prosaic, fairly simple enlargements or explanations of ordinary experiences, and those which are fantastic, of the extraordinary kind. The former, dealing on Tikopia with the coming of ships, the movements of people, the death of workers abroad, the cure of illnesses, are set in the common cultural mold. They represent some of the fruits of speculation about the alternatives of action. Some are fairly simple cases of wrong inference through not having access to all the premises, of which one or more are concealed from superficial observation. An example is the interpretation of a written note as my assistant's intention to stay away for the night.

In them, the instrumental factor⁶ is at a minimum, though it may be present to some degree in giving opportunity to the rumor-monger for ego-assertion. Check is quickly applied and currency is short. In other rumors emotional elements are more marked. There is hope for excitement and novelty, or food supplies, in the rumored arrival of a ship; there is anxiety, as in the rumor of a death. Here, the instrumental factor is more marked and more complex. To some extent it may be argued that the rumor serves as a tension release. By anticipating the event, or possible event, it tones down the emotional excitement and makes it easier to bear. To put speculation into concrete verbal form and give it as a statement, not question, may give subtle reinforcement to the personality by reduction of the unknown to event-shape, to occurrence. Such rumors can be scotched on check—but they tend to revive again.

There is quite a close relation between these

⁶ This is analogous to what Allport and Postman have called "expressive significance," though it is more related to concrete action. (1, p. 169.)

last kinds of rumor, and Tikopia religious speculation involving details of spirit behavior. Rumors that laborers have died overseas, their deaths being reported by radio, are very similar to stories that spirit mediums tell of the end of men who have gone on canoe voyages, their deaths being reported by spirits. Results in social action would be in general the same—preparation of the family concerned for the funeral ceremonies. But there is one significant difference, that with the radio there is opportunity for immediate check. Application by messenger to our house speedily enabled the family concerned to discern rumor from truth, and to avert their mourning. In ordinary Tikopia life, when no anthropologists with radio are resident, there is no such chance of check, and the pronouncement of the spirit medium is normally taken for reality.

In rumors of the fantastic, the untoward event, the extraordinary, other elements are perceptible. In a rumor of a tidal wave, of the suicidal voyage of a man of rank, of the intended violence of chiefs, there are the fruits of speculation of alternatives to action in nature or in the society, as before. But the instrumental factor is much more evident as a move in the field of social action. It may be indirect or direct. A case of its indirect operation is that of a hurricane and tidal-wave rumor. By this, the rumor-monger, it is fairly clear, was making a bid to regain for himself and his family—the leading missionary element—some of the control of public opinion which they had lost. The ostensible subject of the rumor had little or nothing to do with the object desired. But to be in the position of the author of the warning of disaster is tantamount to a claim to the special knowledge, and hence the power, that no one else possesses. A case of rumor with a direct instrumental factor is the story of the meeting of chiefs and executive officials to decide on the expulsion of commoners. This was a false interpretation of circumstances. But its effects were cathartic, the crystallizing of possibilities which had been before men's minds for some time. It was power potential made concrete. Moreover, it could be anticipatory. Rumor may have a proleptic function. The effect of the rumor was to bring out the commoners with arms prepared to defend their lands and homes. This was not necessary. But *if* the thought had

been in the minds of the chiefs that action of expulsion was desirable the effect of the rumor was to trip the wire in advance, to bring the matter to the surface of discussion and obtain reassurances as to its impossibility. To change the metaphor—rumor may be said thus to draw the teeth of intention.

In the case just cited, the instrumental factor was hardly likely to have been consciously employed, after due deliberation. But at a time of crisis a man of rank may allow it to be inferred that he intends to *forau*, to take his canoe and voyage away, to die or survive as the fates decide. This implication is to be distinguished from an announcement of intention, but grades into it. The result is a rumor which spreads through the community, often in exaggerated form. The final outcome of this is to mobilize a force of public opinion against the rumored action. Other men of rank will see that the suicide voyage is not undertaken (4, p. 184). And on the other hand, action is usually taken to correct the position to which the man of rank originally took objection. A further instrumental use of rumor here is as a test of public opinion. By making no public announcement of intention the man has not committed himself, and can withdraw if no counteraction eventuates, i.e., if the public are not seen to be behind him. The "kite-flying" function of rumor can occur in a primitive society, as well as in a sophisticated Western society.

Hence rumor may be said often to have a

positive social function. Its social effects do not necessarily involve any individual or group in scapegoat-suffering. It becomes an organizational mechanism or social instrument in the hands of individuals, seeking to remedy or improve a status position for themselves and the groups they represent. It may tend to maintain rather than destroy a social structure. In this sense the proleptic function which such a rumor may exercise acts as an element of conservatism. But it is a dangerous instrument. Whether its use be conscious or unconscious, rumor in a primitive as in a civilized society is rarely neutral. The fact that it is interest-based almost necessarily gives it destructive potentialities. Whether these are absorbed more easily in a primitive than in a civilized society only further comparative investigation can show.

REFERENCES

1. ALLPORT, G. W., & POSTMAN, L. *The psychology of rumor*. New York: Holt, 1948.
2. FIRTH, R., *We, the Tikopia: A sociological analysis of kinship in primitive Polynesia*. London: Allen & Unwin, 1936.
3. FIRTH, R., *The work of the Gods in Tikopia*. 2 vols. Monographs on Social Anthropology, London School of Economics, Nos. 1, 2. London: Lund Humphries, 1940.
4. FIRTH, R., Authority and public opinion in Tikopia. In: M. Fortes (Ed.), *Social structure: studies presented to A. R. Radcliffe-Brown*. Oxford: Clarendon, 1949.
5. FIRTH, R., Social changes in the Western Pacific. *J. Roy. Soc. Arts*, 1953, 101, 803-816.

Received September 6, 1955.

THE USE OF A PHYSICAL SYMPTOM AS A DEFENSE AGAINST PSYCHOSIS

RICHARD C. COWDEN¹ AND JOHN E. BROWN

VA Hospital, Gulfport, Mississippi

HYPOCHONDRIASIS is one of the first and most obvious symptoms of schizophrenia (1, 3). Through his physical or somatic complaints, the individual may deal with many situations because they are at once a defense against action, a retribution for guilt, and a source of displacement of sensation. For example, ill health (a physical symptom) helps maintain the patient's position in his community. The ailment provides a socially acceptable excuse for not achieving financial gain, social independence, or marital and familial success. "The very doctor who treats him will bear witness to the fact; and for that reason the patient will not need to feel any conscious self-reproaches at using such means to handle problems" (4, p. 44). A hypochondriacal symptom is also a means of maintaining contact with his environment when any real interpersonal relationship is a source of anxiety.

Since anxiety is usually accepted as one of the main stress factors in producing psychotic symptoms, it might be expected that a psychotic's symptoms would decrease if his anxiety could be decreased. This experiment attempted to decrease anxiety in a patient by providing a socially acceptable outlet for it—a physical symptom. Specifically, it attempted to relieve a schizophrenic subject of his psychotic symptoms by replacing them with a physical symptom.

SUBJECT AND METHOD

One patient was carefully selected as the subject for this experiment. The criteria and rationale for selecting the patient were as follows:

1. There should be a unanimous staff diagnosis of schizophrenia. Since the study was concerned with providing remission of primarily schizophrenic symptoms, this point was particularly important.
2. The patient should have had previous unsuccessful treatment with insulin coma therapy

and electric shock therapy. This requirement was a form of protection for the patient and the staff. Since other treatments had failed to aid the patient, there was justification in trying newer, unknown methods so long as they were theoretically reasonable.

3. The patient should have a high level of anxiety. This condition was desirable since studies have shown that conditioning occurs more rapidly in anxious individuals (5, 6).

4. There should be a previous history of some somatic complaint which was not of a nature to require hospitalization and continuous medical treatment. As in hypnosis, so in this study, the experimenter wanted to take advantage of nature in using a response high in the available hierarchy. Thus, an already available symptom was selected rather than an entirely new one. For, as Freud points out, "The production of a conversion symptom is so difficult that an impulsion towards discharge of an unconscious excitation will so far as possible make use of any channel for discharge which may *already be in existence*" (4, p. 53, italics added).

On the basis of these considerations, a patient was selected for the experiment. Following is a summary of his hospitalization and history describing him before the experiment was undertaken:

This 43 year old World War II veteran was admitted to this hospital for the first time in 1953. As a child he was described as being slightly precocious and of having some difficulty in getting along with people. After completing nine grades of school, he began working on construction jobs at an early age. He was inducted into service in 1942, and while overseas as a radio operator on a B-29, he was shot down in 1945. He gave a history of having injured his back falling from a tree after using his parachute. He was a POW for almost a year.

He had had three neuropsychiatric examinations, in 1947, 1950, and 1951, resulting in a diagnosis of anxiety reaction. Hospitalization became necessary in 1953 when he felt that people had it in for him and he began carrying a gun for protection. He had prominent ideas of reference connected with his contracting job. Following a course of ECT, he improved and was discharged with maximum hospital benefits in 1954, diagnosed as schizophrenic reaction, paranoid type.

¹The senior author is now at the VA Hospital, Chillicothe, Ohio.

Two months later it was necessary to hospitalize him again. Following this admission he showed inappropriate affect, was quite grandiose when talking about himself, preoccupied with bodily states, tremulous, and anxious. He was given a course of deep insulin coma therapy, which was completed after 38 treatments and 31 comas. A month after the insulin treatment, examination showed the patient continued to be constricted and rigid, utilizing inadequate compulsive and reaction-formation defenses. Religious ideation and projected guilt were still strong. Nevertheless, he stated that he was ready to leave the hospital. Despite this push to leave and stated optimism, however, his affective structure, with depression and anxiety predominating, seemed unchanged. His defenses were poorly coordinated, and he was extremely anxious and insecure. Projection and autistic distortions of reality were utilized under relatively little stress. He was extremely meticulous in his work, very careful and perfectionistic.

Shortly after this summary was written and he had been given the Rorschach, Bender-Gestalt, H-T-P, a sentence-completion test, and a modified TAT, the attempt to give this man a physical ailment was undertaken. The symptom chosen was a backache. This symptom was selected because the patient's history suggested that it would be more acceptable to him. No actual difficulty existed however.

A "conditioning" method, extending over a period of three months, was followed with all staff members having contact with the patient involved in the actual production of the conversion. X rays were made of his back. Every morning, while making rounds, the ward physician discussed the patient's "back pain" with him, actually starting the experiment by calling the symptom to the patient's attention. At the same time, he was seen in thirty-minute individual psychotherapy sessions twice weekly, during which time the discussion was constantly centered on the selected symptom. The physical therapist provided heat treatments when requested. In occupational therapy he was cautioned not to do any heavy lifting. Social Service helped him in planning for a job with the "ailment" kept always in mind.

RESULTS AND DISCUSSION

At the end of this three-month period, the patient was retested and brought before the staff to be considered for discharge. He told the staff that he now realized his difficulty was caused by a severe back injury which caused him to get anxious and upset. He realized that

his ailment would restrict his activities some but not enough to interfere with his leaving the hospital. The following is a report from a social worker to whom he reported a few days later in another community:

He spoke of his doctor with confidence, at the same time expressing much hostility towards the VA for "having failed to recognize that his trouble had been a fracture of his spine," which occurred when he bailed out of an airplane. He said for eight years he had suffered from his back and all that VA doctors did was to tell him it was "his nerves." In an attempt to get relief, veteran took sedatives (he implied these were narcotics), which, "together with the pain got him in a bad way." Finally, during his present hospitalization, his doctor "listened to him" and gave him heat treatments which relieved him. He now feels he knows what to do should the pain recur.

Further evidence that the patient was using his back pain as a defense was observed in his last testing. Before the physical symptom was developed, he stated, "I think I am better because I have regained contact and favor with God." Three months later, he answered the question, "In what way do you think you are better?" with "... discovery that heat treatment is wonderful for my back. I have developed control over pain." The Rorschach showed little change except for a drop in the *W* responses from 37 per cent to 10 per cent. Some of the other tests gave an indication of a lessening of tension and anxiety. His handwriting was less cramped and tremulous. The Bender showed less erasing, heavy shading, and tremulousness. The H-T-P showed some improvement in that the drawing of a person was more complete. Of interest was the presence of hands on the second drawing whereas none had been drawn the first time. When asked to draw a picture of himself, he was unable to improve his previous performance (2). The sentence-completion test and TAT likewise showed little change in ideation or attitude.

In general, the results of the tests indicate that little real personality change occurred as a result of this unusual treatment. But theoretically none was expected. The treatment made no attempt at changing his ideas or attitudes or basic personality structure; its aim was to provide a new outlet for the expression of his illness which would be socially acceptable. Through this process, a great deal of the anxiety appeared to be

alleviated. The final clinical evaluation made at the hospital by the ward psychiatrist was as follows:

At the time this patient left the hospital on trial visit he had a feeling of well-being, was somewhat euphoric and grandiose, with some expansive plans for employment after leaving the hospital. He tended to be ingratiating, loquacious, but always cooperative. His interpersonal relationships here had shown considerable improvement, and he showed no evidence of being delusional or hallucinated. There was a marked lessening in the manifestation of overt anxiety. He frequently had complaints of pains in his back, and occasionally had mild upper respiratory infection. There was no sensorial disturbance. There was very little insight.

He is now out of the hospital and at last report making a satisfactory adjustment. The last time he left a hospital he had to be returned in less than two months. He has far exceeded this period on this trial; at the origin of this report he had been in the community for nine months.

These results are consistent with Freud's contention that, "...falling ill involves a saving of physical effort; it emerges as being economically the most convenient solution when there is a mental conflict" (4, p. 43). It also suggests that this approach may be of

some practical and economic benefit to hospitals in dealing with certain patients who have failed to improve under more customary treatment. It in no way "cures" the psychosis itself but merely provides the patient with a more socially acceptable manner of manifesting it. Perhaps the germ of a psychosis lies within all individuals; the difference between the patient and nonpatient may not be the actual conflicts and stresses but the means of handling them.

REFERENCES

1. CAMERON, N. *The psychology of behavior disorders*. Boston: Houghton Mifflin, 1947.
2. COWDEN, R. C., DEABLER, H. L., & FEAMSTER, J. H. The prognostic value of the Bender-Gestalt, H-T-P, TAT, and sentence completion test. *J. clin. Psychol.*, 1955, **11**, 271-275.
3. FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.
4. FREUD, S. A case of hysteria. *The complete psychological works*. Vol. 7. London: Hogarth, 1953.
5. SPENCE, K. W., & TAYLOR, JANET A., The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265-272.
6. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-89.

Received September 6, 1955.

CRITIQUE AND NOTES

THREE DETERMINANTS OF THE LEVEL OF ASPIRATION

JAY HANSCHKE AND J. C. GILCHRIST

University of Wisconsin

IN A study of the effects of reference groups on the level of aspiration, Chapman and Volkmann (1) demonstrated that the level of aspiration was inversely related to the prestige value of the reference group. In their study, a "test of literary acquaintance" was given to groups of Ss who were informed of the mean performance of a reference group on the test and who were then asked to give their aspiration-level. Three reference groups of different prestige value were utilized and all were attributed the same performance level. As the prestige value of the reference group increased, Ss aspiration-level decreased, but Ss given the lowest prestige reference group, "an unselected group of WPA workers," though they had the highest aspiration-level, still aspired below the stated performance level of their WPA reference group. The college student Ss had a mean aspiration-level of 33.05 when told the WPA workers averaged 37.2.

Three variables seemed possibly to be affecting the situation: (a) the prestige value of the reference group, (b) the absolute value of the reference score, and (c) the difficulty of the sample questions used in the test directions. The present experiment investigates the effects of these three variables.

METHOD

A pretest was designed to quantify the dimension of sample question difficulty. The Ss were asked to answer a series of multiple-choice questions, and on a five-step rating scale, to state the degree to which they were certain of each of their answers. Acting on the hypothesis that uncertainty of answer is due to difficulty of question, three sets of two sample questions each of low, medium, and high difficulty were established on the basis of their uncertainty scores. The mean uncertainty scores of the sample questions were 1.5, 2.7, and 3.8 for the low, medium, and high difficulty sets respectively. Hereafter, levels of this variable will be represented by Q_1 , Q_2 , and Q_3 in order of increasing difficulty.

The reference groups were chosen from the same, a scholastic, frame of reference. In increasing order of prestige, they were (a) high school seniors, (b) college sophomores, and (c) first year graduate students in psychology. Three reference scores were chosen: low, 54.4%; medium, 69.4%; and high, 84.4%. Following

Chapman and Volkmann, the directions sheets read as follows with appropriate changes for the several conditions:

"On the following pages is a test of general knowledge of psychology. It consists of 50 questions like the following:

Example 1 — The most common type of color blindness involves:

- a. red-green
- b. blue-yellow
- c. red-yellow
- d. blue-green

Example 2 — Which practice requires an M.D.?

- a. psychiatry
- b. psychoanalysis
- c. clinical psychology

"Each question is to be answered by placing the letter corresponding to the one correct answer in the blank before the first word of the question. If you do not know the correct answer, make your best possible guess.

"Your score on the test will be the per cent of questions which you answer correctly. You cannot, therefore, obtain a score higher than 100%, and since there are three or four possible choices for each question, you would probably obtain a score between 25% and 33% by simply guessing.

"This test has been given to a group of first year graduate students in psychology who made an average score of 84.4%.

"Before turning the page, indicate on the line below the per cent correct which you expect to make on this test.

— %

The combinations of the variables in the $3 \times 3 \times 3$ design yielded 27 forms of directions for the test. Ten copies of each form were distributed randomly to 270 students in a large elementary psychology course. The experimental measure utilized was S's level of aspiration from his directions sheet.

RESULTS

Table 1 presents the mean aspiration-level of each of the experimental conditions. The means of the main effects are given in Table 2.

A summary of the analysis of variance is presented in Table 3. It shows the effects of reference group and score to be highly significant, and the effect of question difficulty to be significant not only when isolated, but in the first order interac-

TABLE 1
MEAN ASPIRATION-LEVEL UNDER ALL CONDITIONS

Reference group	Question difficulty	Reference score		
		54.4	69.4	84.4
High school seniors	Q ₁	74.2	81.2	80.4
	Q ₂	76.7	75.0	81.2
	Q ₃	64.4	77.2	54.5
College sophomores	Q ₁	57.8	70.7	80.3
	Q ₂	63.7	68.1	78.5
	Q ₃	57.7	71.2	83.4
First year graduate students	Q ₁	47.0	54.7	56.9
	Q ₂	57.6	53.4	65.2
	Q ₃	36.0	50.9	70.4

TABLE 2
OVER-ALL MEANS

Source	Low	Medium	High
Question difficulty	67.2	69.8	62.9
Reference score	59.4	66.9	72.3
Reference group	73.8	70.1	54.7

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE*

Source	df	MS	F	p <
Reference Group (RG)	2	9,315.6	50.7	.001
Reference Score (RS)	2	3,745.5	20.4	.001
Questions (Q)	2	840.5	4.57	.025
Q × RG	4	585.1	3.18	.025
Q × RS	4	403.0	2.19	—
RG × RS	4	111.1	—	—
Q × RG × RS	8	469.4	2.55	.025
Residual (within)	243	183.8		
Total	269			

* Since the variance was heterogeneous the usually acceptable confidence level was halved. Thus no result was considered significant unless $p < .025$. See Lindquist (2, p. 78-86).

tion with reference group as well as the second order interaction.

DISCUSSION

In a design similar to that of Chapman and Volkmann, the effect of reference group was again found to be a significant anchoring point for Ss' aspiration-level. The effects of two other dimensions of the design, sample question difficulty and reference score, have also been found to have significant anchoring influences which explain the puzzling portion of Chapman and Volkmann's results. In that phase of this study in which the three reference groups were coupled with high reference scores, Ss' aspiration-levels were below that of any prediction based on the anchoring effect of the reference group alone. It is likely that Ss considered performance over this level to be too closely approaching perfect performance and refused to set too difficult levels of aspiration for themselves. As an additional effect of the value of various reference anchors, Ss confronted with the high performance level of a relatively low status reference group seemed to reject the reference group as an anchor and based their aspiration-levels almost exclusively upon the anchor of sample question difficulty. In other words, where the various anchor stimuli are incongruous, Ss tend to reject those which are less bound to the task, in this case reference group and reference score. This last finding raises a significant question: What determines the acceptance of one reference point (e.g., question difficulty) to be accepted as the criterion for judgment while others are rejected? It is to this issue, it would seem, that some research need be directed.

REFERENCES

1. CHAPMAN, D. W., & VOLKMANN, J. A social determinant of the level of aspiration. *J. abnorm. soc. Psychol.*, 1939, **34**, 225-238.
2. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953. Pp. 78-86.

Received June 16, 1955.

ETHNOCENTRISM AND STIMULUS GENERALIZATION¹

FRANKLYN N. ARNHOF

University of Nebraska College of Medicine and the Nebraska Psychiatric Institute

THE phenomenon of stimulus generalization, which has been well documented and experimentally demonstrated (2, 3, 4), has recently come under increased scrutiny with studies relating it to clinical anxiety (5, 13) and various personality correlates (6) including schizophrenic and brain-damaged patient performance (9). Ethnocentrism has recently been explained as an "overgeneralization phenomenon" (10), with at least two studies appearing in the last few years to give some support to this hypothesis (7, 11). The interpretation has not, however, been demonstrated quantitatively. This exploratory study attempts to show that ethnocentrism is related to overgeneralization in thinking, which transcends the immediate attitude of prejudice and has broader implications for theory and research. It was hypothesized that degree of generalization (number of errors on a stimulus-generalization task) is positively related to the subject's scores on a scale of ethnocentrism.

METHOD

The subjects (Ss) were 25 female student nurses who were taking psychiatric training at the Nebraska Psychiatric Institute and who volunteered to serve. Ages ranged from 19 to 22, except for one S who was 27. The majority had spent most of their lives in the more rural areas of the state and had educational backgrounds ranging from high school graduation through two years of college.

The apparatus used for the stimulus-generalization task has been more fully described elsewhere by Brown (4) whose method and procedure were also followed. The apparatus consisted of a six-foot, curved plywood panel with seven 115-v., 7.5-w. frosted lamps mounted horizontally. The lamps were uniformly spaced and were all equidistant from the S who was seated five feet away. A red jewelled pilot lamp, located two inches above the center lamp, served both as a fixation point and ready signal as it was turned on, in random order, 3-5 seconds prior to the lighting of the stimulus light. The S sat with the index finger of the preferred hand holding down a reaction key, and the latency of response was measured to the nearest one-hundredth second by means of a Standard Electric Timer. The experimenter (E) sat behind the panel and by means of switches turned on any of the seven lights following the lighting of the ready signal.

After Ss were comfortably seated at the proper distance, instructions were read stating that this was an experiment on reaction time and that they were to

react as quickly as possible to the lighting of the central lamp by lifting their fingers from the key. They were informed that other lamps would occasionally be lighted to which they were not to respond, and that they should not be upset by errors, but to keep responding as quickly as possible.

Following the reading of the instructions, Ss were given 25 training trials to Lamp 4 (center lamp). After this, without interruption, 84 trials to Lamp 4 were given with 24 trials to test Lamps 1-3 and 5-7 interspersed among them. Each of the three lights to the left and right of the center lamp appeared four times. Six different orders of test trials were used, each beginning with a different test lamp. Four Ss were tested under each of the six orders. As the number of errors (responses to lamps other than Lamp 4) was the primary consideration rather than the shape of the generalization curve, the 25th member of the class was also used as an S under one of the orders.

Each S was identified by a number rather than by name in order to preserve anonymity. After all Ss had performed the stimulus-generalization task, they were seen as a group and given the California Ethnocentrism Scale (1, p. 129), identifying their papers by means of the previously assigned numbers. Upon administration of the E scale it was discovered that one question regarding "zoot suiters" had no meaning to the majority of the group as they did not know to what it referred. Consequently, this item was dropped from the scale because of its obvious invalidity.

RESULTS AND DISCUSSION

In scoring the E scale, 40 points were added to each score to eliminate negative numbers, facilitating treatment of the data. Generalization is here defined as the number of errors made by responding to lamps other than the one to which Ss had been trained to respond. The total number of errors made by each S on the stimulus-generalization task (mean, 6.16; *SD*, 3.21) was then correlated with the individual scores on the E scale (mean, 36.44; *SD*, 9.17.) The obtained value of *r* was 0.44, significant beyond the .05 level for 23 *df* (*p* = .05, *r* = .396). Analysis of reaction times showed no trends or significant differences between Ss.

The results of this study further suggest that the ethnocentric individual makes his judgments and reacts on the basis of overgeneralized thinking. A limiting factor to this interpretation, however, is the well-demonstrated inverse relationship between ethnocentrism and intelligence (1). As it was not possible to obtain intellectual measures on the Ss used in this study, this leaves open the alternative explanation that the obtained relationship between stimulus generalization and ethnocentrism may be due in whole or part to uncontrolled intellectual

¹ The author wishes to thank Dr. Judson S. Brown for his assistance with the details of the apparatus and procedure, and Miss Norma Johnson, nursing instructor, for her help in contacting and arranging for subjects.

variation between Ss, despite the otherwise homogeneity of the group. The relationship between intelligence and stimulus generalization per se remains undemonstrated with the distinct possibility to consider, however, that these two variables may also show an inverse relationship. It is evident, however, that the attitude of the prejudiced individual does have wider implications and correlates as his performance on various cognitive tasks has been shown to differ from that of the nonprejudiced (7, 11).

In discussing the assumptions underlying the phenomenon of empirical stimulus generalization, Brown (4) criticized the explanation offered by Lashley and Wade (8) and Razran (12) who stated that generalization is attributable to a failure to note distinguishing characteristics of the stimulus. In pointing out the circularity of such an explanation, Brown states, "It is evident that if one is to use failure to note to explain empirical generalization, the failure to note must be defined independently of the particular generalization reaction it is supposed to explain" (4, p. 58). In ascribing attributes and characteristics to an individual on the basis of his group affiliation, as does the prejudiced, he is certainly failing to note distinguishing characteristics. Fully realizing the limitations of the present study and the need for more extensive investigation, it is suggested that the relationship expressed here between stimulus generalization and ethnocentrism may supply a necessary independent definition and give strength to the more parsimonious explanation of empirical generalization as being due to a failure to note distinguishing characteristics of the stimulus. More conclusive theorizing along these lines, however, should be delayed until some attempt has been made to investigate the relationship between the various measures of generalization that have been employed in this and previously reported studies. Research along these lines is currently in progress by the author and colleagues.

SUMMARY

This study is an attempt to demonstrate that ethnocentrism is related to an "overgeneralization process" as has been postulated by various authors.

The performance of a group of student nurses on a stimulus-generalization task was correlated with their scores on the California E scale yielding a significant $r = .44$. Some possible theoretical implications are discussed and mention is made of further research which is in progress.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BASS, M. J., & HULL, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17**, 47-65.
3. BROWN, J. S. The generalization of approach responses as a function of stimulus intensity and strength of motivation. *J. comp. Psychol.*, 1942, **33**, 209-226.
4. BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52-61.
5. BUSS, A. H. Stimulus generalization as a function of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 271-273.
6. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561-565.
7. KUTNER, B. Patterns of mental functioning associated with prejudice in children. Unpublished doctor's dissertation, Harvard Univer., 1950.
8. LASHLEY, K. S., & WADE, MARJORIE. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72-87.
9. MEDNICK, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. Unpublished doctor's dissertation, Northwestern Univer., 1954.
10. NUTTIN, J. Personality. *Annu. Rev. Psychol.*, 1955, **6**, 166.
11. O'CONNOR, PATRICIA. Ethnocentrism, "intolerance of ambiguity" and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526-530.
12. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337-365.
13. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35-43.

Received August 1, 1955.

A VALIDITY CHECK ON MMPI SCALES OF RESPONSIBILITY AND DOMINANCE¹

DONALD W. OLMSTED AND ELIO D. MONACHESI

University of Minnesota

GOUGH, McClosky, and Meehl have presented personality scales for Social Responsibility and Dominance which are partly made up of items from the MMPI (1, 2). The scales were derived from a study of high-school and college students rated on responsibility and dominance by peers and teachers who were aided in making their ratings by being presented with a descriptive definition of what the investigators meant by the term.

The 56-item Responsibility scale includes 32 MMPI items, which as a separate scale correlated .84 and .88 with the total scale on two samples reported (2, p. 78). The 60-item Dominance scale includes 28 MMPI items. The correlation between the MMPI Dominance items and the total scale is not reported, but the correlations between the MMPI Dominance items and the criterion ratings (.524 for a college sample, and .597 for a high-school sample) are almost as high as the correlations between the total scale and the criterion ratings (.599 for the college sample, and .687 for the high-school sample) (1, pp. 364-365).

In the case of both scales, the conclusion was drawn that they appeared to have adequate reliability and validity "for the specification of group trends" and differences in most situations (1, p. 366, 2, p. 79). Both scales are mentioned in the 1951 revision of the MMPI Manual (3, pp. 21-22).

PROCEDURES AND RESULTS

The present effort attempts to present evidence concerning the validity and limits of applicability of the MMPI portion of the Responsibility scale in terms of an external criterion and for a quite different sample of subjects. The MMPI portion of the Dominance scale is "tested" less directly.

The subjects are 262 Minneapolis firemen and 25 fire captains, representing about 93 per cent of the firemen on duty on several testing days, (July, 1953) or about 54 per cent of the men in the fire department. The criterion used is the attainment of a captaincy on the basis of the assumption that the captains, on the average, possess "responsibility" to a greater extent than the firemen. With less confidence, the same cri-

terion is adopted concerning "dominance." If these criteria are accepted, the captains should score significantly higher than the firemen on valid responsibility and dominance scales. Data for the subjects are shown in Table 1. Although the captains' mean raw score on the 32 MMPI Responsibility items was slightly higher than the firemen mean, the difference is not statistically significant, and it certainly is not of a predictively useful magnitude. Since the distribution of the scores was somewhat skewed, they were normalized (and a constant added for computational convenience), with a similar result. As Table 1 also shows, the captains had a mean raw score on the 28 MMPI Dominance items slightly lower (nonsignificantly) than had the firemen.

The articles presenting these scales give data concerning the means and standard deviations of the raw scores of the college and high-school students on the MMPI portions of the scales. These are compared with the firemen data in Table 2. The firemen means (including captains) are somewhat higher on both measures. With

TABLE 1
COMPARISON OF MEAN SCORES OF 262 FIREMEN AND 25 FIRE CAPTAINS ON MMPI ITEMS OF SOCIAL RESPONSIBILITY AND DOMINANCE SCALES

Scale	Captains		Firemen		Difference Between Means	SE _d	t	p (One-tail)
	Mean	SD	Mean	SD				
Respon. (raw)	20.16	4.62	19.19	4.39	.97	.9235	1.049	.15
Respon. (normal)	32.20	10.80	30.08	9.82	2.12	2.0734	1.024	.15
Domin. (raw)	15.36	3.86	15.79	3.31	-.43	.7039	-.611	—

TABLE 2
COMPARISON OF MEAN SCORES OF 287 FIREMEN AND 676 STUDENTS ON MMPI ITEMS OF SOCIAL RESPONSIBILITY AND DOMINANCE SCALES

Scale	Firemen		Students*		Difference Between Means	SE _d	t	p (Two-tail)
	Mean	SD	Mean	SD				
Respon. (raw)	19.28	4.41	18.53	4.48	.75	.314	2.37	<.05
Domin. (raw)	15.76	3.36	14.64	3.70	1.12	.254	4.40	<.01

* Student data from (1, p. 365, and 2, p. 79).

¹ This paper reports a portion of the research conducted pursuant to Contract No. N8onr66216 between the Office of Naval Research and the University of Minnesota. The statements expressed are those of the authors, and do not necessarily represent the views of the Office of Naval Research.

respect to Responsibility, the difference between means of .75 units is small enough, and the standard deviations similar enough, to indicate that, whatever this set of items measures, the "tentative norms" established in the original paper may be applicable to adults.

In the research of which this is a portion, it was desired to determine whether the MMPI Responsibility and Dominance score means for the fireman work-groups could be regarded as meaningful measures of group characteristics. Accordingly, the means on these scales of the 40 fire-station work shifts were subjected to analysis of variance tests. The *F* ratio was not significant for either scale, indicating that the group means on these measures differed only randomly. Alternative ways of stating the conclusions would be that the measures do not discriminate among the groups, or that, assuming the scores reflect an aspect of personality, this aspect of personality is not differentially distributed in the groups.

CONCLUSIONS

1. The mean values on the MMPI Responsibility and Dominance items obtained by the present sample of adults were slightly higher than those obtained by the original sample of high-school and college students. The differences, while statistically significant, are not marked, and the tentative norms presented with the scales may be roughly appropriate for adults.

2. The MMPI portion of each scale failed to yield mean values for fireman work-groups (station shifts) which were indicative of differences among the groups in whatever was being measured.

3. Twenty-five Minneapolis fire captains did not differ significantly from 262 Minneapolis firemen in mean scores obtained on the MMPI items of the Social Responsibility and Dominance scales. If the attainment of a captaincy is granted as an appropriate criterion, the data appear to constitute evidence against the validity of at least the MMPI portion of the Responsibility scale. The criterion may be less appropriate with respect to Dominance. Inasmuch as the MMPI Responsibility items were reported in the original study as highly correlated with the total Responsibility scale, some question about the validity of the latter may also be raised.

REFERENCES

1. GOUGH, H. G., McCLOSKEY, H., & MEEHL, P. E. A personality scale for dominance. *J. abnorm. soc. Psychol.*, 1951, 46, 360-366.
2. GOUGH, H. G., McCLOSKEY, H., & MEEHL, P. E. A personality scale for social responsibility. *J. abnorm. soc. Psychol.*, 1952, 47, 73-80.
3. HATHAWAY, S. R., & MCKINLEY, J. C. *Manual, Minnesota Multiphasic Personality Inventory*. (Rev. Ed.) New York: Psychological Corp., 1951.

Received August 8, 1955.

A REPLY TO SCOTT

LEWIS BERNSTEIN¹

Veterans Administration Hospital, Denver, Colorado

SCOTT (4), in a recent article in this Journal, has reported that he was unable to confirm some findings concerning the effects of handling albino rats previously reported by this investigator and co-workers (1, 3). His failure to obtain comparable findings may be a function of differences in design and procedure between his and our experiments.

Scott gentled his animals for three weeks after weaning, in contrast to our reported five weeks of petting (3). After his three-week period he obtained differences in weight gains in favor of his gentled animals. Although these differences were not statistically significant, the trend was in the predicted direction. If Scott had continued the gentling for an additional two weeks he would have replicated at least part of our experiment. This

difference appears important because our data show that the greatest weight gains in our animals occurred during the fourth and fifth weeks of the experimental period.

Scott refers to a personal communication from the author (2) as indicating that there were no differences in water intake between our groups of animals. No mention of water intake was made in that communication. The only paper in which we mention water intake (3) was not referred to by Scott; and in that paper we report that in one of the three experiments the unhandled animals drank significantly more water than the extra-handled animals.

We have reported as one of our major findings that a cessation of handling seriously disrupts the previously handled animals' behavior. For example, "handled animals with whom the relationship was interrupted made significantly more errors than animals with whom the relationship was continued during extinction trials. In fact, the animals

¹ The writer wishes to express his thanks to Drs. John D. Benjamin, Karl F. Muenzinger, and Howard L. Siple for their comments in preparing this communication.

with whom the relationship had been interrupted made more errors than the unhandled animals. In other words, an interrupted relationship produced more errors than a minimal relationship" (1, p. 39). Similarly, in an as yet unpublished study, the animals with whom the relationship was interrupted showed significantly higher death rates following intraperitoneal injection of thiourea than did animals with whom the relationship was continued. With these findings, we would predict that the discontinuance of the handling would disrupt the behavior of previously handled animals on all of Scott's tests; and all of Scott's tests following his three weeks of handling were precisely with animals who were closely analogous to our interrupted relationship group. Scott's gentling stopped when his animals were 44 days old. His test situations were not begun until the animals were 119 days old, and the thiourea injections were not given until they were 136 days old. In other words, Scott was primarily concerned with mature animals, while our effects appeared during, and immediately after, handling of immature rats. On this basis, one can question the validity of Scott's statement that his findings in regard to the fearfulness of his animals failed to confirm those reported by us. We reported greater activity on the part of our handled animals in maze behavior and home-cage activity during the time they were being handled (1, 3). Scott, however, tested fearfulness 75 days after cessation of handling, by placing his

animals in a Miller-Mowrer box for two 3-minute periods.

In the personal communication to Scott, it was stated that the animals were injected "... intraperitoneally with thiourea in dosages of 50 mg. per kilo body weight" (2). It is difficult to understand how injections in the thighs of 100 mg. per 100 gm. body weight (20 times our dosage) can be considered comparable to the injections used in our study.

In summary, Scott reports a failure to confirm our findings on the basis of a study which was not a replication of ours. To the extent that his animals were closely analogous to one of our five groups (the interrupted relationship group), his findings tend to confirm one of our major findings, viz., that an interrupted relationship seriously disrupts the behavior of previously handled animals.

REFERENCES

1. BERNSTEIN, L. A note on Christie's: "Experimental naïvete and experiential naïvete." *Psychol. Bull.*, 1952, **49**, 38-40.
2. BERNSTEIN, L. Personal communication to Jack H. Scott. Jan. 25, 1954.
3. RUEGAMER, W. R., BERNSTEIN, L., & BENJAMIN, J. D. Growth, food utilization, and thyroid activity in the albino rat as a function of extra handling. *Science*, 1954, **120**, 184-185.
4. SCOTT, J. H. Some effects at maturity of gentling, ignoring, or shocking rats during infancy. *J. abnorm. soc. Psychol.*, 1955, **51**, 412-414.

Received March 21, 1956. Prior publication.

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study
in the Behavioral Sciences

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

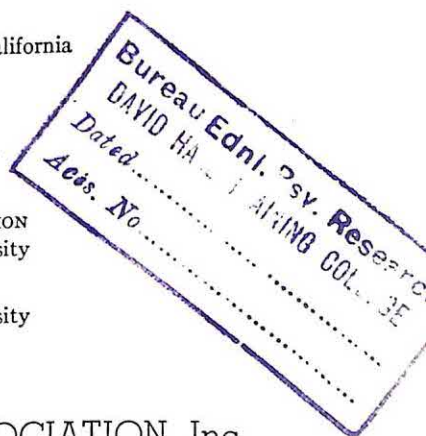
PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6. D. C.

Entered as second-class matter February 15, 1934 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 3440 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc



The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis; those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area, this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 784 pages. The subscription price per year is \$16.00; foreign, \$16.50. Single issues are \$3.00.

Original manuscripts should be sent to the Editor:

DR. M. BREWSTER SMITH
Graduate Department of Psychology
New York University
Washington Square
New York 3, N. Y.

Fifty free offprints are supplied to contributors. Authors of prior publication articles receive no gratis offprints.

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

EMOTIONAL ATTITUDES OF FORMER SOVIET CITIZENS, AS STUDIED BY THE TECHNIQUE OF PROJECTIVE QUESTIONS¹

HELEN BEIER

Boston University

AND EUGENIA HANFMANN

Brandeis University

IN STUDYING the personality patterns of former Soviet citizens we used, among other tests, the projective questions that had been employed in the California studies of authoritarianism (1). These questions were included mainly because, by inquiring about sentiments and cathexes, they elicit expressions of significant attitudes and provide rich material for personality study. Furthermore, since this test had been calibrated as a measure of authoritarianism, we planned to compare the Russian and the American performance in terms of frequency of responses indicating the authoritarian and the equalitarian personality patterns. Because of this double utilization of the data, all test responses were subjected to two separate procedures: they were scored as high, medium, or low in authoritarianism,² and, in an independent operation, they were classified according to their content. The selection of categories for this latter ordering was determined partly by the material itself, partly by our attempting to group the responses in ways that would permit a variety of psychological analyses. The content categories were different for each question and were not intended to reproduce the variables by which the California scoring is guided.

In the present communication we shall report only the results of the content analysis of the Russian and American responses, with the purpose of contributing to the description of the emotional and evaluative attitudes of the Russian subjects. The discussion of the scores of authoritarianism as such must be reserved for a later publication: it could not

be presented without a detailed description of the various alternate ways in which we tried to meet the difficulties involved in applying a scoring system cross-culturally. It should be mentioned, however, that none of the variations of the scoring procedure we used yielded any large differences in the average scores of authoritarianism of the Russian and American samples. In the light of this finding, the differences in the actual content of the responses of the two groups emphasize the fact that both "high" and "low" scores can be given to responses indicating very different kinds of experience and behavior. Thus our results have some bearing on the general problem of subpatterns of authoritarianism and equalitarianism, a topic which will be taken up in the summarizing discussion of the findings.

METHOD

The Russian sample included 39 men and 9 women, almost all of them Great Russians. They were selected for a clinical study more or less at random from a much larger number of former Soviet citizens who were interviewed in Munich in the winter of 1950-51 by the Harvard Project on the Soviet Social System. Of this group of 48 subjects, 30 were brought to Germany during the war: either as war prisoners, or as civilians imported by the Germans to increase their labor force. Their reason for not returning to their native country after the war was fear of the suspicious and punitive attitude of the regime toward those who had spent a long period of time outside the Soviet Union. The remaining 18 subjects (all of them men) found their way to Germany after the war, most of them escaping from the Russian Army of Occupation; some of them were fairly recent arrivals, still very unsettled in their new environment. Younger people, who had grown up under the new regime, predominated in our group: 19 subjects were in their twenties, 19 in their thirties and only 10 older than forty. The middle occupational levels (white collar workers, technicians) predominated, with 23 subjects; the other half of the group were evenly divided between the professional or managerial level, 12, and that of skilled and unskilled laborers, including a few peasants, 13. The distribution of educational levels followed the same pattern of concentra-

¹ The authors are grateful to the Russian Research Center of Harvard University which sponsored this study under a grant from the Human Resources Research Institute of the U. S. Air Force. Drs. A. Inkeles and D. J. Levinson read the manuscript and made many helpful suggestions.

² The scoring in terms of authoritarianism was done by J. Orton and H. Beier with the advice and assistance of D. J. Levinson and Miss L. Heims.

tion in the middle (high school) level, 22, but the college graduates predominated over those who had had only 1-4 years of schooling, 16 and 10 subjects, respectively. About half of the subjects had been affiliated in the Soviet Union either with the Party or with the Communist Youth Organization.

The American group was matched to the Russian group, subject by subject, in terms of age, sex, educational level, and occupational grouping. All subjects came from homes where English was spoken, although, in a few cases, not as the only language; about half of them were born in Massachusetts, the rest being divided between other eastern states, the South and the Middle West. Naturally the two groups could not be matched in terms of their life experiences, recent and past. In both groups the participation in the study was voluntary, and the subjects were mostly recruited through organizations to which they belonged, and were paid for their time. However, for the Russians, the chance to tell the Americans about their lives and their grievances made participation a much more vital experience than it was for the American subjects.

The subjects were given the eight projective questions used in the California studies, and one additional question. (Because of some accidental omissions the number of Russian subjects varies for different questions between 45 and 48.) The questions were asked in the subjects' native language, with probing used when necessary, and with no time pressure. The answers were recorded by the tester. The test was preceded by interviews—very extensive ones in the case of the Russians—so that rapport had been well established. Most of the subjects responded to the questions naturally and willingly.

The content categories for each question were drawn up by the senior author on the basis of a preliminary survey of responses. Since these categories were purely factual, their application demanded no interpretation of responses and presented no particular difficulties; coding as done by the two authors produced very consistent results. If the discussion of a question by a subject contained several separate points they could be entered under different content categories; the term "response" as used by us applies to these units of content. The Russians were more voluble than the Americans and gave a higher total number of responses in 7 out of 9 questions.

The statistical comparison of the two groups was done in terms of number of subjects giving responses in a given content category; the significance of the differences between the groups was tested by means of Zubin's nomographs (3).³ The results of the content analysis will be presented mainly in terms of these comparisons, but some reference will be made also to the distribution of the total responses of each group over the various content categories. Whenever the content categories bear some resemblance to the specific qualitative variables that were mapped out in the California studies, these relationships will be indicated.

³ We wish to thank Dr. T. Alper for her advice and help with the statistical work.

RESULTS

Question 1. All of us are sometimes in a bad mood. What feelings do you think are most unpleasant for you; which of them irritate (upset) you most of all?⁴

Of all nine questions this one elicited from the Russians the largest number of responses referring to vital deprivations and to insecurities inherent in their past and present situation: such references were present in two thirds of the total Russian responses. This high incidence is probably due both to the nature of the question and to its position at the beginning of the series: our subjects utilized this first question for pouring out their most pressing concerns. Responses in one or in both of the closely related categories: "frustration of vital needs" and "negative external conditions" were made by 28 Russians and by 8 Americans.⁵ The category "no prospects in the future" was used by the Russians only (9 subjects**).

Some of the most frequent Russian responses given in these categories refer to lack of work and of vital necessities, to worry about the fate of the relatives who stayed in the USSR, and about one's own uncertain future. This last worry was increased by the fear of the Soviets' invading Germany which was widespread in Europe at the time of the study: such an event would spell disaster to the nonreturnees. Following are some illustrative quotations from the records:

I am most disturbed when there is nothing to eat, no money to buy clothing, particularly if one of the children is sick. When all are in good health you can bear these lacks more easily. Right now I ought to buy shoes for my boy—he has not been able to go out of the house for a week. But if we buy clothing we need, not enough would be left for food.

All the moral and physical sufferings that result from our abnormal life: absence of regular work which would provide personal satisfaction and the material basis for living. This problem is with us always: this futile chase after work.

At home it was fear of getting involved with the

⁴ In most of the questions we simplified the original formulations to make them more understandable to the subjects of lower educational level in both national groups; consequently the version given here is somewhat different from that used in the California studies.

⁵ This difference is significant at the .01 level of probability. Such differences will be indicated in the text by double asterisks, those significant at the .05 level by single asterisks.

NKVD—how to escape them. Now I am disturbed by the feeling that war might start any day, that I might wake up and find the Red Army here; and this is a fear not just for myself.

The American discussions contain no comparable responses. Those few answers that have been categorized under "negative external conditions" refer not to vital but to trivial or even freakish situations: "I am upset when people smoke in my car;" "When somebody pours cold water on my back while I am asleep." Mentions of this kind of "external condition" are interpreted in the California studies as projections of neurotic anxiety into some irrelevant situations; this interpretation is much less applicable to the responses of the Russian group which, in terms of the California variables, seem rather to indicate concern with "threatening or non-supporting environment."

The next largest category for the Russians is that of "interpersonal situations" (24 Russians, 16 Americans*). It comprises two distinct subcategories which show a clear-cut differentiation between the two groups. The majority of the Russian responses refer to personal disharmony as an *interactive relationship* in which the subject feels himself participating, be it in the active or in the passive role (23 Russians, 6 Americans**). The responses of the Americans refer more often to moods and traits of *other people* (grouchy people, overbearing people, irresponsible drivers), with the subject himself as an outsider, suffering from these offensive traits (2 Russians, 11 Americans**).

In terms of California variables the Russian responses may be viewed as indicating "libid-inized interpersonal relationships." This interpretation is also supported by the frequency with which the family is mentioned, whether in terms of longing, of concern for their well-being, or of being upset by discord (12 Russians, 1 American**). Furthermore the Russian responses also contain references to injustice and to the sufferings of other people, particularly of the Russian people under the Soviet regime, and to the lack of understanding of their situation by the West: "When one does not understand the Russians, when one confuses Russia with the Soviet Union—I will argue till I drop uncon-

scious that the Russian people are not, as a whole, infected with Communism." These semi-ideological responses based on identification with a group have no counterpart in the American discussions (13 Russians**).

The category most frequently used by the Americans is that of "own moods." While the Russians use it with almost equal frequency (17 Russians, 21 Americans), the specific content of the responses of the two groups shows as much difference as in the case of "external conditions" and of "interpersonal situations." The Russians talk about longing, loneliness, fear, and moral sufferings; they never mention boredom, irritability, and restless tension which function prominently in the American responses. These objectless moods form a transition to the category of "bodily conditions" which is used by the Americans only and which contains complaints about poor health, sleeplessness, and—most frequently—fatigue (10 Americans**). In terms of California variables such references to bodily conditions are assumed to express "ego-alien trends," such as passivity, anxiety, and hostility. On the other hand, the Americans also exceed the Russians, though not significantly so, in the category "frustration of achievement" (10 Americans, 5 Russians), a type of response which is supposed to reflect conscious conflict and self-criticism.

Question 2. All of us have some desires which we try to suppress. Which desires and feelings do you find most difficult to control, most difficult to suppress?

The Russians made some references to specific conditions of their present and past, talking, e.g., about how difficult it was to suppress the wish to fight, to protest against the Soviet regime, or to banish painful thoughts about the relatives left behind. However, such references were much less frequent than in Question 1. Apart from this special category and from the Russians' greater difficulty in understanding the question, the distribution of the major content was fairly similar in the two groups.

In each group one third of the subjects made references to difficulties in the "control of primitive impulses." Within this category the Russians mentioned aggression and drinking with equal frequency, while for the Americans the concern with hostile impulses pre-

dominated over all others (7 Russians, 13 Americans). In terms of the California variables this might mean the prevalence of "non-focal aggression" in the American responses, and a greater frequency of the "incidental pleasures" variable in those of the Russians. However, the difference between the two groups was significant only in the subcategory of drinking, which was used only by the Russians (7 Russians*).

The Americans used two categories that were absent in the Russian material. The first comprised the wishes "to quit," "to go off," "to do what I want to do and not what I must": to get away from the obligatory routine (7 Americans*). Such responses are interpreted in the California studies as indicative of "ego-alien passivity." The second category referred to the difficulty the subjects had in suppressing other disturbing feelings and moods (5 Americans*). The third minor category, which comprised "wishes for fame and material possessions," showed no differentiation between the two groups (7 Russians, 7 Americans).

Question 3. Whom do you consider to be really great people? What kind of people do you admire?

The bulk of the responses of both groups consisted of names of admired people, given with or without elaboration. They fall into two major categories: "artists and scientists" (named by 34 Russians and 28 Americans), and "social leaders" (23 Russians, 28 Americans). A smaller proportion of responses of both groups was given in terms of admired personal traits.

The subcategories of "social leaders" yielded no significant differences between the two groups, though the Americans tended to favor liberal leaders, such as Lincoln and F. D. Roosevelt (13 Russians, 20 Americans); the Russians, while also paying tribute to those "who worked and lived for the good of the people," gave no less attention to names that connote national power, including the military (15 Russians, 8 Americans). It should be noted, however, that in the latter subcategory the Russian list of choices was headed by Peter the Great, who represents progress and reform as much as national power; among the purely military leaders the

Russians gave prominence to General Kutuzov who is credited with defeating Napoleon and who stands for victorious defense of the motherland rather than for conquest.

Pronounced differences between the two groups appear in the subcategories of "artists and scientists." Scientists are mentioned by them with about equal frequency (7 Russians, 10 Americans), but while the Americans give only a slightly greater prominence to artists of all descriptions, the Russians mention the latter with much greater frequency (33 Russians, 13 Americans**). When the artists are subdivided into writers (including poets), musicians, and painters, the Americans are shown to distribute their choices among these groups evenly, while the Russians favor the writers (31 Russians, 7 Americans**), naming musicians much less frequently and painters hardly at all. The difference between the two groups with regard to admiration for writers appears even more striking if we take into account that the Russians gave 86 single responses in this category and mentioned 34 different names, while the corresponding frequencies for the Americans are 10 and 8. Prominent among the writers and poets mentioned by the Russians were Pushkin, Tolstoi, Dostoevsky, Lermontov, Yessenin; of the foreign writers Shakespeare, Victor Hugo, Jack London, and Mark Twain were mentioned more than once.

The Russians' great admiration for writers is a continued cultural tradition; it may indicate the high value they place on expression and representation of human emotions, as compared with objective understanding and with practical action. The greater stress placed on these latter by Americans is reflected in the relatively greater weight they give to science and to social leadership, in their occasionally naming businessmen and sportsmen whom the Russians do not mention at all, and in the nature of the character qualities they consider admirable. Within this latter category they give primacy to traits related to practical achievement, such as strength, capability, success, while the Russians stress more inward and "moral" attributes, such as sincerity and kindness. However, the only difference within these minor categories that reaches statistical significance is the greater frequency with which

American subjects express admiration for their own relatives (2 Russians, 9 Americans*).

Question 4. Nearly every person says at times to himself: If this goes on I shall go out of my mind. What things can make people go out of their minds, lose their senses?

In this question the frequency rank order of the three major categories is the same for both groups, although the actual frequencies and the specific content of responses are different in some cases. For both groups the main cause of mental disturbance lies in internal psychological conditions, in "emotional and mental stress" (26 Russians, 30 Americans). In the category "external conditions" the Russians predominate over the Americans (20 Russians, 13 Americans), though not significantly so. This relationship is reversed in "disturbed personal relations" (4 Russians, 14 Americans*).

The differences between the responses of the two groups which appear in the subcategories are similar to those observed in Question 1. Of the Russian responses coded under "external conditions," three-fourths refer to conditions of extreme pressures and threats which are not mentioned by the Americans (14 Russians**); on the other hand, the Russians never mention minor environmental disturbances such as noises, disorder, traffic jams, referred to in more than half of the American responses (7 Americans**). The few Russian remarks coded under "disturbed personal relations" refer to interpersonal discord, while the Americans take a more passive attitude in their mentions of being persecuted, criticized, exposed to a nagging wife or to a drunken husband, or of being isolated from "normal relationships." It is interesting to note that while in Question 1 the Russians stressed the importance of personal relations for their happiness or unhappiness more than did the Americans, in thinking about the causes of a more catastrophic breakdown the latter show more concern than the former about the disastrous consequences of conflict and isolation.

In the leading category of "emotional and mental stress," the Russians talk mostly about stirring traumatic experiences which may overpower the person, such as terror, deep grief, or unhappy love (26 Russians, 11

Americans**); only the Russians occasionally blame insanity on lack of control of emotions or of impulses (5 Russians*). The Americans ascribe mental illness predominantly to conflicts and frustrations created by internal obstacles, by "blocks within oneself," "no outlets for your emotions": a type of explanation never given by the Russians (20 Americans**). This difference in the concept of mental stress can be traced in part to linguistic factors and to relatively superficial cultural influences. Since psychoanalytically oriented theories have been condemned by the "party line," even the educated Russians have not been exposed to them as the American public has been. Terms like "repression" or "defenses" are absent from their vocabulary and from their thinking; the popular term "frustration" which was the one most frequently used by the Americans in the category of emotional stress has no exact counterpart in Russian. However, the difference in the conception of emotional causes of insanity might also reflect some real differences of psychological functioning. The apprehension of the Russians that they will be disorganized by overstrong emotions might be a consequence of their greater emotional abandon and impulsiveness; such experiences could make them stress the necessity of deliberate control, of holding impulses and emotions in check. In terms of California variables, most of their responses would fall under the category of "too much inner life." The Americans, on the other hand, sound as if they were greatly hemmed in by automatic defenses against strong emotional experiences to the extent that they possess an introspective orientation, and, in the vocabulary for describing these "inner psychological states" of conflict and tension, they express their apprehension of the consequences of this situation by ascribing insanity to frustration and repression.

Question 5. What do you think is the worst crime a person can commit?

The main content category which comprises one-half of the responses of each group is "murder." This category was subdivided, as in the California studies, into plain unqualified references to murder, and into responses that, by bringing in the motives for the act, or by some other qualifications and

elaborations, go beyond the reference to the external act as such. The first subcategory showed no difference between the two groups (20 Russians, 20 Americans); however, more Russians than Americans gave responses in the qualified subcategory, such as "murder committed for financial gain," or "murder by the government" (16 Russians, 5 Americans**).

References to stealing and deceit were made by Russian subjects only (9 Russians**). References to "offenses against oneself and one's values" were made only by Americans (5 Americans*). Betrayal of friends or country, cruelty, and sexual offenses were mentioned by both groups with equally low frequency.

Question 6. All of us get sometimes into situations when we feel very much ashamed and embarrassed. What experiences make you feel like sinking through the floor?

The bulk (three-fourths) of each group's responses to this question fall into three categories: "violation of moral values," "inadequacy," and "violation of conventions." The first refers to failure to live up to one's obligations, promises, or to moral standards of behavior; the formulation often implies that the person was ashamed of what he had done, and not merely humiliated by exposure. The second category refers to experiences of personal inadequacy and inability, either in work or in interpersonal situations. Responses in the third category show a preoccupation with conventional rules of behavior and etiquette and a fear of appearing ridiculous: having one's clothes in disorder while in public, making an involuntary *faux pas* in using a foreign language. This category may be considered as the opposite of the first, because the emphasis here is shifted from essential values to superficialities of behavior, and from one's own concern with one's action to the unfavorable reaction of others. The category of "inadequacy" occupies a middle position between the two. Some responses in this category reflect the person's own concern with achievement and personal adequacy, but more often than in the case of "moral values" the cause of painful embarrassment lies in the exposure of one's inadequacies to others.

The distribution of responses over these categories is very different for the two groups.

While the Americans use the three categories with approximately equal frequency, the Russians refer most frequently to moral values (30 Russians, 13 Americans**), and very rarely to conventions (6 Russians, 16 Americans*). This distribution suggests that the Russians are less afraid of exposure of their personal weaknesses to others, and experience less social anxiety than do the Americans. No difference was found between the two groups in the frequency of the use of the category "inadequacy" (13 Russians, 17 Americans), but its specific content tends to be different. The Americans talk almost exclusively about intellectual inferiority, mistakes, or failure to work, being proven wrong or criticized in front of others; the DPs' feelings of personal inadequacy are often aroused by having no work and no earnings, by being forced to ask for help. In "violation of moral values" both groups equally emphasize lying and dishonesty, with a minor focus on "hurting people" in other ways, but while the Americans express these concerns in a rather impersonal or self-contained way ("failure to live up to promises," "to discover I was offensive"), the Russians often make the situation concrete by referring to specific acts and to particular people involved; they feel regret, e.g., if they had let down a friend, or had been rude to an old person. As in Question 5, occasional references to stealing are made only by the Russian subjects.

Question 7. Suppose you knew you had only six months to live, but during that time you could do whatever you liked—how would you spend your time?

This question invites the subjects to explore in what direction his wishes would take him if all obstacles to his actions, including their long-term consequences, were removed. The question also evokes reactions to imminent death. In both respects there are pronounced differences between the two groups.

References to "enjoyment, pleasure" form the main category for both groups (28 Russians, 29 Americans); however, the distribution of responses within this category is quite different. In the Russian group references to frankly *sensual pleasures*, such as eating, drinking, and sex, outweigh almost two to one the mentions of such pleasures as travel, enjoyment of arts and sports (sensual pleasures: 22 Russians, 8 Americans**). In terms of

California categories the predominant Russian responses belong under "open sensuality and active pleasure." Conversely among the Americans, references to what one subject termed "*the finer things of life*" predominate over the cruder pleasures in the relationship of three to one, "travel" being their most frequent single response (11 Russians, 25 Americans**). These responses have a marked affinity to the "incidental, dilute pleasures" of the California classification.

Other differences are less pronounced. The Russians give more responses in the category of "religious and moral preoccupations," expressing these concerns either in pure form or as an integral part of some conflict about pleasure seeking or impulse gratification (16 Russians, 6 Americans*). They also tend to dwell more on the "affiliative ties," if this category is made to include not only the wish to spend time with family and friends, but also the wish to see Russia once more (14 Russians, 9 Americans). Neither group expresses any desire to utilize the time left for contributing to social welfare, to mankind in general; however, some of the Russians wish to relieve the suffering of the Russian people, e.g., by committing a decisive terroristic act, such as killing Stalin. The Americans emphasize personal achievement slightly more than do the Russians, particularly in the form of a wish to continue and complete some piece of work.

In expressing their reaction to imminent death, subjects in both groups comment that this knowledge might deprive them of any enjoyment and enterprise. However, while the Russians occasionally talk about "going crazy" or "sitting and crying," of committing suicide, the Americans often deny the emotional upset explicitly or by implication, maintaining that they would carry on as usual; some of them express the wish to keep their fate secret from their families and friends.

Question 8. We have the feeling of awe when something seems to us remarkable, beautiful, or important. What things or events inspire such feelings in you?

Even though an elucidation of the word "awe" was included in the question, this term was frequently not understood, or misunderstood, particularly by the Russians. They often gave it the meaning of "admiration" and occasionally the meaning of "pleasure" or

"delight." Responses referring to such events as being given a chance to emigrate are caused by this misunderstanding which possibly also accounts for some of the differences in the responses of the two groups.

In the American group the leading category is that of "individual achievement" (10 Russians, 28 Americans**). In more than half of the instances the Americans specify this achievement as scientific or artistic, but—unlike the Russians—they also make references to outstanding performance in technology and in sports (1 Russian, 10 Americans*), while the Russians refer to arts and sciences almost exclusively. The second place for the Americans is occupied by "nature," which includes also references to life processes (6 Russians, 16 Americans*). In the Russian group the leading category is the "social-interpersonal": the subjects feel moved by warm personal relations and altruistic acts, as well as by events of wider social significance such as the triumph of freedom (15 Russians, 7 Americans*). Responses expressing "national pride" were equally infrequent in both groups (5 Russians, 7 Americans).

Within the categories used by both groups, the formulations of the Americans tend to be more schematic and conventional than those of the Russians, indicating a less intense and personal mode of experience. Thus in the category of "nature and life" most of the American responses are so brief and stereotyped that they must be described as "dilute experiences" in terms of the California categories. The Russian responses are much more concrete and detailed, more suffused with personal meaning and emotion, and consequently may be considered as representing "intense nature experiences."

Question 9. Suppose you had a child, and you knew that you might wish three things for him and that your wishes would be fulfilled—what would you wish for?⁶

Apart from the fact that the Americans gave many more general unspecified responses, such as "happiness," (5 Russians, 22 Americans**), the wishes for the child were fairly similar in the two groups. The largest category

⁶ This question has been used in T. Dembo's study, "Investigation of Concrete Value Systems," U. S. Public Health Service, Institute of Mental Health. Final Report (Unpublished).

contained wishes for a "successful life" or for a life lived under favorable conditions (30 Russians, 34 Americans); in this category the Russians stressed particularly the importance of education and of specialized training as prerequisites for the "good life" (17 Russians, 9 Americans). The Americans, on the other hand, placed a greater emphasis on wealth, on financial status (8 Russians, 16 Americans*).

While both groups wished for good health for the child (20 Russians, 27 Americans), the Russians emphasized no less strongly mental qualities (18 Russians, 12 Americans) and moral or character traits (23 Russians, 11 Americans**) to which the Americans paid less attention. In speaking of mental qualities, the latter mentioned intelligence almost exclusively; the Russians divided their wishes between intelligence and special talents, particularly the artistic ones (9 Russians, 2 Americans*). The Americans showed slightly more concern than the Russians with the child's social relations and also made occasional references to "full life," to "secure and harmonious development," which had no counterpart in the Russian responses. The Russians occasionally expressed the wish that the child should grow up to love Russia, and be able to apply his talents in Russia.

DISCUSSION

In reviewing the trends revealed in the responses to the nine questions, we shall first organize the discussion in terms of the positive and negative values held by the two groups, and then consider the implications of our findings for some aspects of the theory of the "authoritarian personality." We shall use the term "value" in a wide sense, including not only explicit evaluations, but also consistently positive or negative emotional reactions of a more immediate kind, as they are expressed both in the content and in the formulation of responses. It is well to remind the reader at this point that the generalizations that follow are based merely on the differential frequencies of responses in some of the content categories and that many of these differences are small, or occur within minor, infrequently used categories. In several questions there is a considerable similarity between the two groups, at least in the distribution of the major content categories. Furthermore, some of the dif-

ferences that do appear are obviously created or enhanced by the drastic differences in their life situations both in the present and in the recent past, and cannot be ascribed exclusively to differences in permanently held values. On the other side of the ledger, it is to be noted that many of the differences observed are not explicable in situational terms alone, and that several of the intergroup differences appear quite consistently from question to question, even though they may reach the level of statistical significance only in some of them.

In reviewing those questions that elicit expressions of *positive values* (Questions 3, 7, 8, 9), we find that two major values are more prominent in the Russian records than in the American ones. The first one is the value of emotional experiences and of frank sensual pleasures. The Russians' positive acceptance of these experiences is expressed both in the content of their responses and in the language they use, which is more emotionally expressive and more concretely descriptive of sensual impressions than that used by the American subjects. Even more pronounced is the Russians' high valuation of interpersonal relationships, their positive and active acceptance of interaction and of belongingness with others. This attitude is expressed in most of the questions, both in an immediate emotional fashion and on an ideological level. The other two features more prominent in the Russian than in the American responses—moral values and patriotic feelings—also have strong interpersonal implications for our subjects. Patriotism is often expressed in the context of wishing to help the Russian people; moral values, such as sincerity and honesty, are seen as prerequisites of friendship.

The value which the Russians emphasize less strongly than do the Americans is first and foremost that of individual achievement; the latter mention it more frequently than do the former in practically all contexts that evoke discussions of this topic. Within this category what seems distinctively American is the high valuation of rational knowledge and of organizational and physical achievement. A second positive category in which the Americans have the advantage is an infrequent one but it has practically no Russian representation. This category includes responses emphasizing independence and self-expression of the

individual and a protest against their violation; the concept of "crime against oneself" (e.g., "deceiving oneself") also belongs to this category, which could be termed "integrity of the individual."

In response to inquiries about *negative values* and sources of negative experiences (Questions 1, 2, 4, 5, 6), the Russians much more often than the Americans see causes of unhappiness in severely depriving conditions of life, conditions that obstruct satisfaction of vital needs or threaten one's life or safety. The other two categories in which their responses by and large exceed those of the Americans refer to disturbance of interpersonal relationships and to violation of moral values. These three categories, which can be viewed as the obverse of the Russian positive values, belong to three different spheres: external, interpersonal, and ethical. Yet all of these categories refer to vital conditions of human existence and imply the person's essential interrelatedness with his environment. Unhappiness is seen by the Russians as resulting from the disruption of this relationship either by the "environment" (depriving conditions) or by the person (moral violation), or by both (interpersonal discord).

In contrast to this relative homogeneity, the negative conditions that are mentioned by the comparison group more often than by the Russians fall into two major categories which seem to be quite different in their psychological meaning. On the one hand, consistent with their positive values of individual achievement and individual integrity, the Americans locate causes of unhappiness within the person himself: in his conflicts, inhibitions, in his feelings of inadequacy and failure. These responses presuppose a self-reflective, self-critical attitude which is less prominent among the Russians. On the other hand, we find numerous references to inessential, often trivial conditions which have little relation to basic human needs, and which are located "outside": in other people, in things, in the body. Annoyances produced by noises, bossy people, insomnia, breaches of etiquette, or impulsive wishes to "go off" are likely to be merely symptoms of disturbances caused by unconscious conflicts. Such responses indicate a defensive alienation from one's feelings which is the opposite of insight. Yet both categories

have the common feature of implying the person's separateness rather than close relatedness with his environment.

Thus a review of the distinctive values of the two groups suggests that the Russians are more firmly and securely integrated with their environment than are the Americans, who are more keenly aware of the individual's separateness and isolation. Because the most significant part of our environment is people, it is understandable that the most pronounced differences between the two groups appear in the interpersonal area.

Considering our findings in terms of the subvariables of the equalitarian and authoritarian personality patterns, we arrive at the following formulations. The equalitarian patterns of the Russians are based to a greater extent than those of the Americans on the area of *interpersonal relationships*, while those of the Americans stem from their concern with the *rights and achievements of the individual*. Within the area of inner experience, the Russians' equalitarian personality structure is attested to by their *acceptance of impulses and emotions*, and by moral self-reproach, while the corresponding American pattern finds expression in a more rational and self-critical *introspective attitude* and in feelings of inadequacy and rejection. The authoritarian patterns of the two groups also show differences but along different lines. If one were to score all Russian responses according to the specific rules of the California study which have been worked out for the American groups, a greater proportion of the Russian than of the American "high" scores would be earned through "moralizing" discussions and through expressions of patriotic sentiments—i.e., through their *ideological attitudes*. The high scores of the Americans are more often founded on *defensive personality aspects*, such as alienation from oneself and from others and displacement of emotions into irrelevant situations.

These generalizations must be qualified by some further observations which are pertinent to the status of ideological attitudes as indicators of personality patterns. The wording and the context of our subjects' responses suggest that certain explicit evaluations do not have the same significance for the two national groups, and this impression is confirmed by their differential correlations with the over-all

scores of authoritarianism. Thus the Russians' traditional admiration for great writers does not seem to be diagnostic of personal equalitarianism to the same degree as it is for the Americans, nor do their feelings for their country appear to indicate ethnocentric authoritarian patriotism. Insofar as the choice of particular values and opinions, even when guided by personal motives, depends also on their prevalence and meaning in the cultural environment, such differences are to be expected. The more direct dependence of ideological attitudes on the environment makes them less valid as indicators of either "high" or "low" patterns than are the more deeply ingrained personal traits, particularly when subjects of different cultural backgrounds are being compared. Since our delineation of the difference in the authoritarian patterns of the two groups is based in part on the Russians' ideological pronouncements, it appears less valid than the generalizations that concern the intergroup differences in equalitarianism.⁷

⁷ The formulations concerning the equalitarian patterns are also borne out by the evidence of some other tests we used in the study. In comparing the subjects' scores on different tests J. Orton found this evidence to be particularly clear-cut on the point of "interpersonal" vs. individual approach. One of the items in the Episodes Test (2) depicted a conflict between the individual and the group of which he is a member. In discussing this situation, the Russians in general identified with the group more strongly than with the individual, but this tendency was even stronger among those whose scores of authoritarianism on the Projective Questions were low than among the "highs." The Americans as a group displayed a stronger identification with the individual, and here the "lows" did so to a greater extent than the "highs." To understand this paradox, one must take into account that the Russians identified with the group in a very positive, participating way, while the majority of the Americans described the group as a coercive agent to be either fought with or yielded to, i.e., in terms of an irrational authority.

Some evidence on the "emotional" aspect of the "low" Russian pattern was obtained from the technique of "short answer items," used with the Russian group only. The responses of the Russians to these questions, most of which pertained to various areas of Soviet life, were scored in terms of the four scales of Flattery, Distortion, Rejection of the basic Soviet institutions, and general Anti-Soviet sentiment. Flattery (i.e., opinions favoring the Americans as, for example, over the British) was higher among the "highs" than among the "lows": the former are more eager to please the momentary authority. For the rest of the variables the relationship was reversed. The "lows" not only

By comparison with the attitudes of Americans, the equalitarian traits of the Russians appear to be more pronounced in the interpersonal than in the "individual" area, and in the emotional rather than in the intellectual sphere. What implications do these findings have for the general theory of authoritarianism? Primarily they serve to confirm the existence of subvariants, at least within the equalitarian pattern, and to identify the two variants that are distinctive of the two national groups: one centered on interpersonal relatedness and immediate emotional awareness, the other on the integrity of the individual and on rational self-reflection. Though these two patterns reflect different value-orientations, and possibly also different levels of sophistication, within the theoretical framework of the California studies they may be considered as equally expressive of genuine equalitarianism. It is a task for future studies to verify or disprove the existence of these and of other possible patterns by a systematic investigation of the correlations between the many specific traits that are considered as diagnostic of equalitarianism and of authoritarianism.

SUMMARY

The projective questions used in the California studies of authoritarianism were administered to a group of former Soviet citizens and to a comparison group of Americans. The responses were coded according to their content. The findings are discussed in terms of the emotional and evaluative attitudes prevalent in each group, and their implications

expressed a much stronger anti-Soviet sentiment than the "highs," but also went much further in rejecting such widely accepted institutions as the state's ownership of industry, and even in denying such recognized achievements of the regime as increase in literacy and in production of farm equipment (Distortion). It would seem that respect for either factual objectivity or for achievement as such is not a component of the Russian "low" pattern. Although these subjects would not falsify their evaluation of the American merits and failings while talking to us, for them to pass an objective judgment on the instrumental achievements of the Soviets would violate the emotional truth in a more important way than would detached objectivity. This interpretation is borne out by their frequent alternative reaction to inquiries about the Soviet achievements: "Yes, but at what price—and to what end!"

for the theory of authoritarian personality are pointed out.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. HANFMANN, EUGENIA, & GETZELS, J. W. Inter-
personal attitudes of former Soviet citizens, as studied by a semi-projective method. *Psychol. Mongr.*, 1955, **69**, No. 4 (Whole No. 389).
3. ZUBIN, J. Nomographs for determining the significance of the differences between the frequencies of events in two contrasted series or groups. *J. Amer. statist. Ass.*, 1939, **34**, 539-544.

Received July 25, 1955.

VOLUNTEERING AS AN AVOIDANCE ACT

ROBERT R. BLAKE, HOWARD BERKOWITZ, ROY Q. BELLAMY
AND JANE SRYGLEY MOUTON

The University of Texas

VOLUNTEERING is a dynamic social process. The amount of volunteering can be increased or decreased by changing the structure of the force field acting on the person in the specific situation. Some of the factors determining the resultant force toward or away from accepting a volunteering request already are known. For example, volunteering can be increased by others accepting (8, 9) and decreased by others refusing the request (8). The strength of the request also influences acceptance since frequency varies directly with the compellingness of the induction (7). Still another factor is whether or not acceptance is in public or private, with more volunteering under private conditions of induction (9). Although clear relationships between personality and volunteering have not been demonstrated (6), it can be presumed that personal variance is a factor in the resultant force toward or away from volunteering.

The hypothesis tested in this study is that changes in the frequency of volunteering in response to a standard request are associated with variations in the attractiveness of possible alternative actions. The prediction is that the more attractive the alternative the lower the frequency, and, conversely, the lesser its attractiveness the higher the volunteering rate. The prediction is the same whether made on the basis of a Lewinian analysis (3), adaptation-level theory (1, 2), or from approach-avoidance concepts (4). Since a previous experiment has demonstrated differential acceptance rates for requests under public and private conditions (9), the hypothesis tested here is evaluated for both conditions of responding.

METHOD

Subjects and Arrangements

Subjects included 362 students in undergraduate classes at The University of Texas. Six classes served under the private condition. Slips were distributed which students were requested to sign; they were asked to check either: "I wish to participate in the experiment," or "I do not wish to participate in the experiment." Three additional classes served under the public condition where students indicated acceptance by raising their hands.

Variations in the Attractiveness to the Alternative to Volunteering Condition 1 (Attractiveness associated with not volunteering).

A positive expectancy for the alternative to volunteering was created by the instructor's informing the class he would be called out during the hour. He indicated that the class would be dismissed for the remainder of the period. The experimenter then entered to solicit volunteers. After a few seconds' hesitation, the instructor agreed to excuse from class those who volunteered.

Condition 2 (Control condition). Shortly after the instructor had begun to lecture, the experimenter entered and solicited volunteers as described above. The instructor agreed to excuse from the remainder of the lecture those who volunteered.

Condition 3 (Volunteering as an avoidance act). A negative expectancy was created by the instructor's announcing a "pop quiz." While the questions were being written on the blackboard, the requester entered to solicit volunteers. The instructor explained that a test was being given. After a few seconds' hesitation, he agreed to excuse from taking the test those who volunteered.

Request for Volunteers

After receiving permission to solicit volunteers, the requester made the following appeal. "We are conducting a psychological experiment and need some volunteers to participate this hour. It will take no longer than this class period." For the private condition he said, "Will each of you sign your name to the slips that are being passed around, indicating whether or not you will participate? Check one of the two statements on the slip." For the public condition the following statement was substituted for the last sentence, "Will those of you willing to participate raise your hands?" A full explanation was provided after completion of the study, since subjects were unaware that the arrangements were part of an experiment.

RESULTS

The data for the three conditions for both "private" and "public" volunteering are presented in Table 1. Consistent with prediction, the lowest rate was obtained when the alternative was attractive (Condition 1). Being dismissed produced significantly less volunteering than the opposite condition involving the unattractive alternative of a "pop quiz" and less, but not significantly so, than the control condition. The prediction of a higher rate for Condition 3 also is confirmed since the unattractive choice of taking a "pop quiz" is associated

TABLE 1

FREQUENCY OF VOLUNTEERING AS A FUNCTION OF BACKGROUND CONDITIONS

Condition*	N	Number of Volunteers	Percentage by Condition
Private			
1	67	18	26.8
2	83	38	45.7
3	85	84	98.8
Public			
1	34	4	11.0
2	39	12	30.7
3	54	54	100.0

* In Condition 1, attractiveness is associated with not volunteering; in Condition 3, volunteering avoids an unpleasant situation; Condition 2 is neutral in these respects.

TABLE 2

CHI-SQUARE VALUES FOR VOLUNTEERING UNDER DIFFERENT BACKGROUND CONDITIONS

Condition*	χ^2	df	p
Private 1 \times 2 \times 3	22.83	2	.01
1 \times 2	2.53	1	—
2 \times 3	9.82	1	.01
1 \times 3	18.83	1	.01
Public 1 \times 2 \times 3	24.80	2	.01
1 \times 2	2.79	1	—
2 \times 3	9.65	1	.01
1 \times 3	16.36	2	.01
Private \times Public	.54	1	—

* In Condition 1, attractiveness is associated with not volunteering; in Condition 3, volunteering avoids an unpleasant situation; Condition 2 is neutral in these respects.

with significantly more volunteering than is found for the control condition or for Condition 1. Here volunteering is an avoidance act. The invitation is accepted in order to leave an undesirable situation. Frequency for the control condition gives a rate intermediate between the extremes. The significance of the differences between conditions are shown in the χ^2 s in Table 2.

The same relationships were obtained for both public and private conditions, though volunteering in private for Conditions 1 and 2, is greater than volunteering in public. The latter trend though not significant is similar to the one reported by Schacter and Hall (9). The findings gain further support from a study where personal independence in the Asch-type situation was maintained more frequently when the anonymity of the respondent was insured (5).

DISCUSSION

The rate of volunteering in any concrete situation, then, is a function of the social field structure. The strength and direction of the resultant force determines the frequency with which acceptance occurs. The attractiveness of the requested action relative to that of available alternatives is a critical determinant of the resultant force. The general principle is that the rate can be changed by increasing or decreasing the desirability of the task, by increasing or decreasing either the attractiveness of alternative actions, or by both in varying proportions.

The findings can be evaluated from the standpoint that the subjects reported experiencing the decision to volunteer as a matter of personal preference. Participants saw little or no connection between the experimentally created expectancies for the class hour and the request for volunteers. They felt they were acting on personal preference, yet from the standpoint of the experimental design, volunteering was regulated by the relative attractiveness of the experimentally created alternatives. The difference between the situation experienced by the subject and that which occurred provides a basis for assigning meaning to the term "volunteering." The subject feeling that action is volunteered, or governed solely by personal preference, occurs whenever the respondent sees no causal relationship between the various factors that are involved in his decision. He feels he has made a personal decision concerning participation, yet from a systematic point the resultant force that determines his decision is based on a predictable weighting of all relevant factors operating in the situation. Thus, decisions that in personal experience seem to be free, independent, and personal are, from another standpoint, determined by the psychological properties of the force field within which the action occurs.

Examples may be given applying the findings reported here. Why do people volunteer for work such as in the American Red Cross, the American Cancer Society or the Girl Scouts of America? The decision is determined by the attractiveness of the work, relative to possible alternative activities. Military recruitment through voluntary enlistment also can be viewed in the same manner. The rate of enlistment is high when civilian employment is diffi-

cult to procure, as in times of depression, and low during prosperous times, when alternative civilian employment is more desirable.

SUMMARY

Volunteering as a function of the relation between the attractiveness of the requested action and that of the alternative to volunteering was studied in the present experiment. Three hundred sixty-two subjects reacted under public or private conditions to a standard volunteering request. Attractiveness was varied so that under one set of conditions volunteering would be more desirable than the alternative, while under another it would be relatively more undesirable. Under the control condition attractiveness of the alternative was not altered. Results support the predictions. Volunteering was significantly more frequent when the alternative was less attractive than the act and less frequent when the alternative was more attractive than the requested action. Frequency under the control condition was intermediate between the other two. The conception of volunteering as a dynamic social process resulting from the structure of the force field was discussed.

REFERENCES

1. BLAKE, R. R., & MOUTON, JANE S. The study of social conduct within the framework of adaptation-level theory. In M. Sherif & M. O. Wilson (Ed.) *Emerging problems in social psychology*. Third conf. in Social Psychol. Norman: Univer. of Oklahoma, in press.
2. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, 55, 297-313.
3. LEWIN, K. *Field theory in social science*. New York: Harper, 1951.
4. MILLER, N. E. Comments on theoretical models illustrated by the development of a theory of conflict. *J. Pers.*, 1951, 20, 82-100.
5. MOUTON, JANE S., BLAKE, R. R., & OLMSTEAD, J. A. The relationship between yielding and disclosing personality identity. *J. Pers.*, 1956, 24, 339-347.
6. ROSEN, E. Differences between volunteers and non-volunteers for psychological studies. *J. appl. Psychol.*, 1951, 35, 185-193.
7. ROSENBAUM, M. The effect of stimulus and background factors on the volunteering response. Unpublished doctor's dissertation, Univer. of Texas, 1954.
8. ROSENBAUM, M., & BLAKE, R. R. Volunteering as a function of field structure. *J. abnorm. soc. Psychol.*, 1955, 60, 193-196.
9. SCHACHTER, S., & HALL, R. Group-derived restraints and audience persuasion. *Hum. Relat.*, 1952, 5, 397-406.

Received September 22, 1955.

THE STRUCTURE OF SOCIAL ATTITUDES IN LEBANON

E. TERRY PROTHRO AND JACK D. KEEHN

American University of Beirut

ONE of the authors of *The Authoritarian Personality* has observed (12, p. 73) that in spite of the enormous number of studies inspired by publication of that volume, the identification of the authoritarian syndrome is still far from established. In examining data on this subject, Christie (3, p. 140) was reminded of Augustine's comment about time: "For so it is, oh Lord my God, I measure it, but what it is that I measure I do not know."

In contrast to the uncertainty of these writers is the opinion of Eysenck (5, p. 152), who states quite simply that "the F-scale is essentially a measure of tough-mindedness." By "tough-mindedness" he means performance on a scale developed by himself (4) and co-workers (5, p. 276). The T scale is unifactorial and some idea of its importance may be gained from the assertion that it, together with the unifactorial R scale designed to measure radicalism-conservatism, determine almost all correlations between all attitudes in England, the United States, Sweden, Germany, and other countries having similar forms of social organization (5, pp. 265-266). If Eysenck's view can be accepted, then it becomes clear that social psychologists have been neglecting the key to many of their most difficult problems—including that of the authoritarian syndrome.

The evidence cited for the proposition that "the F-scale is essentially a measure of tough-mindedness" is far from conclusive. Coulter, in an unpublished doctoral dissertation (5, pp. 152-153) found a correlation of .43 between the T scale and the F scale for a group of 83 soldiers, a correlation of .62 for a group of "male Fascists," and a correlation of .63 for a group of "male Communists." The prohibitive value of this evidence should be clear to anyone familiar with the vast number of scales and indices which have moderate positive correlations with the F scale (10, p. 23).

Coulter also found that her Communists scored higher on the F scale than did her unselected soldiers, and Eysenck seems to feel that this reveals new information about the F scale. This datum is sufficiently surprising

to call for closer examination, for other evidence (11, pp. 24 ff.; 3, p. 131 ff.) points in the opposite direction, but in any event it bears only indirectly on the identity of F and T. Consequently it appears that the suggestion that the F scale and the T scale are measuring essentially the same thing must be viewed as a bold but unproved assertion.

Keehn (7) administered Eysenck's public-opinion inventory (4) in Lebanon with a view to determining whether the R and T dimensions would emerge. He did find two orthogonal factors which he labeled "Arab nationalism" and "toughmindedness-tendermindedness." He concluded that his results were compatible with those of Eysenck, but that Eysenck's R factor was a special case of the general dimension which might be called "the major political issue of the culture under investigation." It should be noted, however, that for most items Keehn's two factors account for less than 20 per cent of the total item variance.

It has been shown that the F scale is a useful and meaningful instrument for measuring authoritarianism in Lebanon (9) in spite of cultural differences between California and Lebanon. Some question was raised as to the meaning of authoritarianism, however, when it was found that there was no correlation in Lebanon between the F scale and the California scale of politico-economic conservatism. Moreover, a follow-up study (8) revealed that there was no correlation between authoritarianism as shown by scores on the F scale and ethnocentrism as shown by performance on scales of social distance. In interpreting this lack of relationship it must be noted that there is no evidence that either the PEC scale or the social-distance scale taps a meaningful dimension of social attitudes in the Near East.

To explore the meaning of authoritarianism in such a setting it would seem appropriate to employ items which have emerged from factorial studies performed in the culture. Performance of Lebanese subjects on the F scale was therefore compared with their response to items representing each of the two factors found by Keehn, toughmindedness-tendermindedness and "Arab nationalism."

TABLE 1
FACTOR SATURATIONS OF ITEMS SUBJECTED
TO ANALYSIS

Item*	Unrotated			Rotated		
	F _I	F _{II}	F _{III}	F _I ''	F _{II} ''	F _{III} ''
F6	51	42	-35	73	06	-10
F8	-05	26	35	00	-07	42
F9	37	-32	12	-03	49	-05
F12	07	11	12	08	03	15
F13	78	-10	-15	46	61	-17
F16	45	18	-05	42	23	05
F18	59	37	05	58	27	23
F19	70	-28	12	19	74	-03
F21	35	36	-36	60	-03	-13
F25	40	16	12	30	26	18
F26	41	-05	-32	34	24	-30
F29	23	17	-30	38	-01	-17
F31	20	34	-15	41	-07	03
F33	06	-23	-57	11	00	-60
F34	54	-29	-13	18	53	-25
F35	25	37	13	36	04	29
F38	44	27	40	30	32	47
F39	47	-23	17	06	54	03
F42	34	50	26	45	08	46
F44	68	-23	42	10	78	25
T9	29	-13	14	-06	22	24
T19	22	26	-15	-36	02	00
T20	36	22	34	24	26	40
T36	-17	13	16	-08	-17	20
N8	-21	45	-26	27	-49	00
N13	33	-25	07	00	41	-07
N18	-54	54	-30	16	-80	00
N39	40	-26	15	01	48	00

* Both here and in the text the item numbers are those of the sources (1, pp. 255-257; 4, p. 566) rather than of our questionnaire. The letters F, T, and N refer to fascism, toughmindedness-tendermindedness, and nationalism respectively. See footnote 2 re item N13.

METHOD

Subjects were 100 students in General Education and Educational Psychology courses at the American University of Beirut. Of these, 72 were male and 28 female. With respect to nationality, 54 were Lebanese and 46 from nearby Arab states. During regular class periods they were given Eysenck's public-opinion inventory as adapted by Keehn (7) for use in the Near East together with an F scale of 27 items.¹ The format and instructions for the two scales followed the standard form of the original scales. Subjects worked anonymously. They had no difficulty in completing the two inventories in a single class session.

A centroid analysis was then conducted on items from the F scale and the Eysenck scale. The item inter-correlations for the analysis were tetrachoric correla-

tions computed by use of the diagrams of Chesire *et al.* (2). Responses were dichotomized into "Agree" or "Disagree," with "Neutral" responses grouped with the smaller category in each instance. Because of the unreliability of such correlations when the dichotomy is near the tail of the distribution, items were omitted from the analysis when fewer than 20 per cent or more than 80 per cent of the subjects agreed with the item. Seven F-scale items were omitted for this reason. The factor analysis was performed on 28 items, of which 20 were from the F scale, 4 were from Keehn's toughminded-tenderminded scale, and 4 from the "nationalism" scale. The toughminded-tenderminded and "nationalism" items were selected on a basis of their high loadings in Keehn's previous study (7).

RESULTS

Factor saturations for the 28 items are shown in Table 1. Orthogonal rotations toward a simple structure were carried out graphically. There appear to be two significant factors running through the items analyzed. Analysis was terminated after extraction of three factors because the third factor was not statistically significant according to the methods of Tucker and Humphrey as described by Fruchter (6, pp. 77 ff.), and because only one of the second factor residual correlations reached the .01 level of significance whereas about four would have been expected by chance.

All but two of the F-scale items have positive loadings on Factor I''. Thus the belief that there is a general factor in those items is supported by our data. At the same time, however, it should be noted that some of the items have small loadings on this factor and large loadings on a second factor. The items which seem to be the best measures of Factor I'', in order of size of loadings are:

F6 human nature being what it is, there will always be war and conflict

F21 young people sometimes get rebellious ideas, but . . .

F18 . . . people move around and mix together so much, a person has to protect himself against catching a disease or infection from them

F13 what youth needs most is strict discipline . . . and the will to work and fight . . .

F42 no . . . decent person could ever think of hurting a . . . relative

F16 some people are born with an urge to jump from high places

F31 people are prying into matters that should remain personal . . .

F29 astrology can explain a lot of things

F35 sex life of the old Greeks and Romans was tame compared to . . . this country . . .

T19 major questions of national policy should be decided by reference to majority opinion (negative)

¹ All items of Form 45 were used except for items number 22, 30, and 43 (1, page 260). Two of these items referred to events at the close of World War II and one was a terse folk saying.

These items do not fit readily into any single cluster which has yet been found from analyses of the F scale. It seems to the writers that all of the items reflect what has been called a "low opinion of human nature" and of human effort. Item F42 might be interpreted as the traditional Near Eastern reliance on one's own in-group as the principal defense in a world where human nature is untrustworthy. We have therefore named this factor "cynicism," although admittedly other names and other interpretations are possible.

Factor II" seems to be the factor which Keehn called "Arab nationalism." The four items which he used to define this dimension have significant loading on Factor II". The items which our analysis shows to have loadings of .35 or more on this factor, in order of magnitude of loading, are:

[N18 Jews are as valuable, honest and public-spirited citizens as any other group (negative)

F44 nobody ever learned anything... except through suffering

F19 an insult to our honor should always be punished

F13 what youth needs most is strict discipline... and the will to work and fight...

F39 homosexuals... ought to be severely punished

F34 most of our social problems would be solved if we could somehow get rid of the immoral, crooked and feeble-minded people

F9 when a person has a problem... it is best for him not to think... but to keep busy...

N8 in the interests of peace, we must give up part of our national sovereignty (negative)

N39 the Japanese are by nature a cruel people

N13 refusal to fight for ones' country on religious grounds should be a punishable offense²

Some of these items might be thought of as characteristic of the views held by Arab nationalists, but items F39 and N39 would be hard to fit into such a concept. On the other hand, items with heavy loadings, such as F44, F19, F13, and F39 clearly reflect authoritarian aggression. It would be easier, then, to think of Factor II" as one of authoritarian aggression or punitive outlook. Item N18 might appear on first reading to bespeak nothing but ethnocentrism. However, when the Arab-Israeli conflict and the problem of the Jewish minority in Lebanon are considered in connection with the item, it seems possible that punitiveness is involved in the response.

² This item was edited by Keehn (7) and the wording therefore differs from that which appeared in Eysenck's inventory (4, p. 566).

It is interesting to note that the authors of the F scale listed (1, p. 255) eight items which they had selected, on a priori grounds, as typifying authoritarian aggression. Six of these items were entered into our analysis, and four of them (F19, F13, F39, F34) had substantial loadings on Factor II". A fifth item (F12) had virtually no loading, but it referred to "bad manners" and "breeding" and the item may simply be inappropriate in this culture. The sixth item (F25) had a loading of .26. The item referred to "...rape and attacks on children... such criminals ought to be publicly whipped, or worse." Such an item seems to reflect punitiveness in a high degree, and the low loading might raise some question about the identification of Factor II". From discussion with local students and social scientists, the authors have discovered that public punishment is associated in the minds of many Arabs with the Ottoman occupation and with foreign mandatory powers. That is, such punishment is not thought of as an act on the part of the ingroup. Moreover, it has been pointed out to us that public punishment for sex crimes would make a public disgrace of an issue which might otherwise be settled by punitive action on the part of the family. It is possible, then, that these connotations of the phraseology of the item explain why it did not have a higher loading on the punitive factor.

Item F13 is listed as having a moderate loading on both cynicism and punitiveness. It is our interpretation that this item reflects a low opinion of the value of creative effort on the part of Arab youth and at the same time taps the punitive factor by its reference to fighting for family and country.

If we accept "punitiveness" rather than "nationalism" as the label for the second dimension, this implies that it is a dimension of personality and not of the political sphere. This point of view leaves open the possibility that some nationalists might hope to achieve national unity and independence without reliance on authoritarian aggression. It further implies that, contrary to Keehn's earlier suggestion, this factor cannot be linked with Eysenck's radical-conservative factor under the general heading of "major political issue of the culture under investigation."

The items which Keehn found to be measures of toughmindedness-tendermindedness in

Lebanon do not seem to be effective indicators of either Factor I" or II". The identification of authoritarianism with toughmindedness does not hold with our sample. Indeed, there is for this sample a correlation of only 0.19 between an F scale composed of the entire 27 items and a T scale composed of the 10 items which Keehn had found to be the best measures of T in Lebanon. For the F scale, the mean was 4.7 and the standard deviation .71. For the T scale, the mean was 1.86 and the standard deviation 5.11, where the highest possible score was 20 and the lowest possible score was -20.

SUMMARY

Previous studies with the F scale in the Near East had raised some question about the factors being measured by the scale. In order to gain more information on authoritarianism in this culture, an analysis was made of items from the F scale together with items from a scale of Eysenck's which seemed to measure toughmindedness and a political dimension. It was found that there was a general factor running through almost all of the F scale items, and this factor was tentatively labeled "cynicism." The "toughminded-tenderminded" items were poor measures of this factor. A second factor was extracted, which seemed to be the same as that previously identified as a political dimension. Our data suggested that this factor might better be interpreted on a personality dimension, which we tentatively labeled as "punitiveness" or "authoritarian aggression."

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CHESHIRE, L., SAFFIR, M., & THURSTONE, L. L. *Computing diagrams for the tetrachoric correlation coefficient*. Chicago: Univer. of Chicago Bookstore, 1933.
3. CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.). *Studies in the scope and method of "The Authoritarian Personality"*. Glencoe, Ill.: Free Press, 1954. Pp. 123-196.
4. EYSENCK, H. J. Primary social attitudes: a comparison of attitude patterns in England, Germany, and Sweden. *J. abnorm. soc. Psychol.*, 1953, **48**, 563-568.
5. EYSENCK, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.
6. FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
7. KEEHN, J. D. An examination of the two-factor theory of social attitudes in a Near Eastern culture. *J. soc. Psychol.*, 1955, **42**, 13-20.
8. PROTHRO, E. T. Generalized ethnic attitudes in the Arab Near East. *Sociol. soc. Res.*, 1953, **37**, 375-379.
9. PROTHRO, E. T., & MELIKIAN, L. H. The California Public Opinion Scale in an authoritarian culture. *Publ. Opin. Quart.*, 1953, **17**, 353-362.
10. SANFORD, N. Recent developments in connection with the investigation of the authoritarian personality. *Sociol. Rev.*, 1954, **2**, 11-33.
11. SHILS, E. A. Authoritarianism: "right" and "left." In R. Christie & Marie Jahoda (Eds.). *Studies in the scope and method of "The Authoritarian Personality"*. Glencoe, Ill.: Free Press, 1954. Pp. 24-49.
12. WEBSTER, H., SANFORD, N., & FREEDMAN, M. A new instrument for studying authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73-84.

Received September 27, 1955.

RORSCHACH AND TAT INDICES OF HOMOSEXUALITY IN OVERT HOMOSEXUALS, NEUROTICS, AND NORMAL MALES¹

ANTHONY DAVIDS

Brown University and Emma Pendleton Bradley Home

MARK JOELSON AND CHARLES MCARTHUR

Harvard University

THERE has always been much concern regarding the nature, cause, and prevalence of homosexual behavior. Although the extent of this behavior in contemporary America can be estimated only crudely, reports by social scientists suggest that homosexual activity of some kind, prepubertal, adolescent, or adult, is far more prevalent than is generally realized (3, 5).

Although the diagnosis of homosexuality is not difficult to make for those individuals who do not attempt to hide their abnormality, there are many persons who recognize and perhaps act on desires which are more than latent, yet who deny them. The attitudes of society and demands of social living cause these individuals to retreat alone into the shadows with their problem. Thus, the diagnosis of homosexuality by ordinary direct methods of assessment is often difficult or impossible. The development of projective techniques, however, opened up new avenues for diagnosis and analysis of psychological dynamics. The present investigation, using the Rorschach test and the Thematic Apperception Test (TAT), is an attempt to contribute to further understanding and clarification of the problem of homosexuality in young adult males.

More specifically, this study, employing a group of overt homosexuals, a group of patients suffering from other neurotic difficulties, and a control group of normal subjects (Ss), will:

1. Compare the Rorschach protocols of these groups in regard to homosexual signs derived from the psychological literature.

2. Compare the TAT protocols of these groups in regard to homosexual signs derived from the psychological literature.

3. Compare and correlate the results on both tests within each of the groups.

4. Suggest for further research any new signs uncovered in this study that show a significant difference between the groups.

METHOD

Subjects

Homosexual group. The overtly homosexual group was composed of 20 men who, at the time of testing, were students attending large Eastern universities. They were all concerned about their problem, wanted to talk to the psychologist about it, and were aware that the psychologist knew they were overt homosexuals. These men were all tested by the same psychologist in his office in a university clinic setting. Their intelligence ranged from high average to superior, and all came from average or better socioeconomic backgrounds.

Neurotic group. The neurotic group was composed of 20 male students who were attending a large Eastern university and were being seen for psychological diagnosis and psychiatric treatment at the university's hygiene department. These Ss were suffering from neurotic emotional disturbances. If diagnostic labels were to be used, the group would consist mainly of anxiety states, obsessive-compulsives, hysterics, and depressives. All of these patients were of high average or superior intelligence, and came from average or better socioeconomic backgrounds. The examining psychologist, who also tested the homosexuals, affirmed that there were no known homosexuals in this group. In the present study, these emotionally disturbed Ss, who were also tested under conditions conducive to frankness, constitute a control group. That is, this group was introduced to ensure that differences found between the homosexuals and the normal control Ss, described below, were not a function of a generally disturbed state but were reflections of specifically homosexual disturbance.

Normal group. The normal Ss were 20 male students attending a large Eastern university and, at the time, participating in an intensive investigation of personality being conducted by a team of trained investigators. These Ss were selected by an experienced clinical psychologist who tried to secure a sample of the college population containing a diversity of interests, abilities, and personalities. The men were above average in intelligence and came from average or better socioeconomic backgrounds. They had all volunteered to serve as Ss for the research project, participated in it over a two-year period, were guaranteed anonymity, and were told that the data would be used solely for purposes of sci-

¹ This study was facilitated by the Laboratory of Social Relations and the Department of Hygiene at Harvard University, and by research grant M-700 from the National Institute of Mental Health, Public Health Service.

entific research. Therefore, it is reasonable to expect their responses to be honest and valid.

From these descriptions, it is evident that the three groups were essentially comparable in respect to most significant variables. For purposes of the present study, it is important to point out that there were no significant differences between the groups in regard to college major. Each group appeared to contain a fairly representative sampling of the various fields of concentration, with few Ss in each group majoring in subjects indicative of strong aesthetic interests. It can be stated

with considerable certainty that the groups did not differ significantly in regard to literary or aesthetic interests. The one major variable that differentiated these groups was the overtness of homosexual practices in one, whereas the other two, as far as it could be ascertained, contained no overt homosexuals.

The Rorschach Test and Thematic Apperception Test

Administration. Standard instructions and procedures for administration of the TAT (9) and the Ror-

TABLE 1
RORSCHACH SIGNS OF HOMOSEXUALITY*

Sign	Card	Location	Content	Examples
1	I	Whole or part of whole	Mask or human or animal face	"Theatrical mask" "Fox, leering"
2	I	Lower center detail	Male or muscular female torso	"Japanese man" "Woman athlete"
3	II	Lower center detail	Crab or crablike animal	"Deep-sea crab" "Cooked lobster"
4	III	Whole or part of whole	Humans; with sex confused	"Woman above, man below," "Man, but with breasts"
5	III	Whole or part of whole	Humans; with sex uncertain	"Two sexless people" "Could be either"
6	III	Whole or part of whole	Animal or animal like (dehumanized)	"Featherless chickens" "Ostrich-like natives"
7	IV	Whole or part of whole	Human or animal; contorted, monstrous, or threatening	"Man looking back through legs," "Horrid beast"
8	V	Whole, part of whole, or center detail	Human or humanized animal	"Woman dressed as bat" "Bugs Bunny"
9	VI	Center or top detail	Object; with implication of cleavage	"Ship with foaming wake" "Oil well, drilling"
10	VII	Whole, part of whole, or top detail	Human; female with derogatory specification	"Two old hags" "Women—yakety-yak"
11	VIII	Lateral detail	Animal; several incongruous ones or one with incongruous parts	"Rat or a lion or something," "Bear with a rat's head"
12	IX	Upper lateral detail	Human; dehumanized	"Witch-like creature" "Monster with guns for arms"
13	X	Top center detail	Animal; attacking or fighting over central object	"Beavers gnawing a tree," "Ants arguing over this"
14	X	All or upper half of pink plus center blue	Human; with blue as oral specification	"Ectoplasm coming from mouth," "Irishman with growth on lips"
15	Any card, I-X	Any location	Human or animal	"Decayed teeth" "Open mouth showing tonsils"
16	Any card, I-X	Any location	Human or animal anal detail or specification	"Woman with huge ass" "A bloody rectum"
17	Any card, I-X	Any location	Human or animal described as "back to back"	"Two Negroes with their backs pressed together" "Two sheep back to back"
18	Any card, I-X	Any location	Human object or architecture with religious specification	"A priest's cross" "Two men praying"
19	Any card, I-X	Any location	Male or female genitalia	"Part of a man; his nuts," "A woman's private parts"
20	Any card, I-X	Any location	Feminine clothing	"A woman's fur jacket" "Panties"

* Adapted from Wheeler (13, pp. 104-105).

schach (6) were followed. The complete Rorschach test (10 cards) was administered to all Ss in the three groups. However, there was some variability as to which TAT cards were administered to different Ss as well as differences in the manner of recording the stories. The TAT stories of the homosexual and neurotic groups were recorded, using a form of shorthand and later typed in detail in the attempt to secure verbatim protocols. The stories of the normal group were electrically recorded on Dictaphone belts and were transcribed verbatim. As for cards administered, there was considerable inconsistency in the homosexual group. The TAT records for this group ranged from one card (the test being abandoned when the S became disturbed) to 15 cards. It was possible, however, to form subsamples, within the homosexual group, of those who had all responded to the same stimulus cards. Fifteen of the 20 homosexuals responded to Cards 4, 6, 7, 10, and 18 of the male series. Also, five of the homosexuals were administered Cards 1, 3, 4, 6, 10, 11, 13, 14, and 18, which permitted further comparisons within this more limited sample. Thus, a few of the "signs" derived from the TAT literature could not be adequately tested because the required card had not been administered to the homosexual group. The neurotic group responded to 10 TAT cards: 1, 3, 4, 6, 7, 8, 13, 14, 17, and 18; the normal group composed stories for the entire set of 20 cards in the male series of the TAT. In spite of this lack of consistency of cards administered to the Ss, the available TAT data proved quite satisfactory for purposes of the present investigation.

The Rorschach signs and their scoring. In considering which Rorschach indices of homosexuality, taken from the psychological literature, offered the most promise, it was decided that the list compiled by Wheeler (13), drawn in part from three earlier studies (2, 4, 7), represented the most complete and specific scheme available. Although the Wheeler scheme has been utilized elsewhere, it had never been put to a crucial test with a group composed entirely of overt homosexuals. Detailed description of these signs, their derivations, and rationale, is presented in Wheeler's report (13). An abbreviated description of the 20 signs is presented in Table 1. It is apparent that the majority of these signs involve a minimum of ambiguity in their scoring. The data were scored "blind," so that the experimenter did not know if he was scoring the protocol of a homosexual or a control S. The scorer went through each protocol scoring the presence or absence of each sign and recording the total number of homosexual signs in the entire protocol.

The TAT signs and their scoring. There is a dearth of organized schemes containing TAT indices of homosexuality. Tejesy (11), however, has presented a list of empirical generalizations concerning "typically homosexual" responses as suggested by scattered statements in the TAT literature. In a later paper, Lindzey, Tejesy, Davids, and Heinemann (8) modified and lengthened the list of empirical generalizations and quoted additional authorities for some of the indices previously reported. Detailed descriptions of the 10 homosexual signs, presented briefly in Table 2, with a discussion of their source and rationale, is contained in these two manuscripts (8, 11). The present investigators, after a careful survey of the literature, were unable to find any published reports of the experimental test of the presence of these signs in protocols of overt

homosexuals. The TAT records for the three groups were scored "blind," the scorer recording the presence or absence of each sign and the total number of homosexual signs in the protocol.

Reliability of scoring. An independent investigator scored the Rorschach signs in 10 of the homosexual protocols and 10 of the normal group protocols. The interrater agreement of scoring was 83 per cent; the product-moment correlation between the number of signs scored by each rater was .92. Another independent judge scored the TAT signs in 10 of the homosexual protocols and 10 of the normal group protocols. The interrater agreement of scoring was 96 per cent, and the product-moment correlation between the number of signs scored by each rater was .85.

Hypotheses

Hypothesis I. The homosexual group gives a significantly greater mean number of proposed homosexual Rorschach signs than does the neurotic group or the normal group.

Hypothesis II. The homosexual group gives a significantly greater mean number of proposed homosexual TAT signs than does the neurotic group or the normal group.

Hypothesis III. Within each group there is a significant positive correlation between Rorschach and TAT signs of homosexuality.

RESULTS

Rorschach Test

In support of Hypothesis I, the homosexual group gave a significantly greater mean number of Rorschach signs proposed as indices of homosexuality than did either of the control groups. As shown in Table 3, the homosexual group gave a mean of 5.2 signs, while the normal control group gave a mean of 2.8, and the neurotic group a mean of 2.3 signs. For both comparisons, the difference

TABLE 2
TAT SIGNS OF HOMOSEXUALITY

Sign	Card	Content
1	9	Card makes subject uncomfortable
2	10	Subject thinks that scene is a sham, that there is really no reciprocal affection
3	12	Stories in which the young man has been hypnotized by the older man
4	18	Stories in which the hero has been attacked from the rear or pulled to the rear
5	Any card	A high degree of feminine identification
6	Any card	A recurrent shift of identification
7	Any card	Misrecognitions of sex of TAT figures
8	Any card	Derogatory attitude toward marriage and the opposite sex, including stories of marital discord
9	Any card	Stories in which men kill women
10	Any card	Sexual, genital references

TABLE 3
RORSCHACH SIGNS OF HOMOSEXUALITY AND PRODUCTIVITY OF RESPONSES

	Homosexual Group (N = 20)		Neurotic Group (N = 20)		Normal Group (N = 20)		Homosexual vs. Neurotic		Homosexual vs. Normal		Neurotic vs. Normal	
	Mean	Vari- ance	Mean	Vari- ance	Mean	Vari- ance	F	t	F	t	F	t
Number of signs	5.2	6.7	2.3	5.4	2.8	1.9	1.24	3.76**	3.54**	3.66**	2.84*	.79
Number of responses	67.9	1860.8	45.0	798.2	50.6	385.3	2.33**	1.99*	4.83**	1.63	2.07	.73
% signs per response	9.3	22.5	5.3	30.6	6.3	17.6	1.36	2.45*	1.28	2.13*	1.74	.65

* Significant at the .05 level.
** Significant at the .01 level.

between the homosexuals and the nonhomosexual Ss is significant well beyond the .01 level. However, the means of the two nonhomosexual groups do not differ significantly.

In order to determine whether the significant difference in signs might have been caused, in part, by a difference in total responses, the difference between the mean number of responses given by each group was measured. The results presented in Table 3 show that the mean number of responses in the homosexual group is significantly higher than the mean in the neurotic group, but not significantly higher than the mean in the normal group when evaluated according to a two-tailed test of significance. A test of the difference between the neurotic and normal groups, as shown in Table 3, was not significant.

Since the homosexual group gave more signs and also produced more responses than the control groups, the mean percentage of signs given in proportion to total number of responses was computed for each of the groups. It was found, as shown in Table 3, that a significantly greater percentage of the homosexual group responses were homosexual signs, and that the mean percentage of homosexual signs did not differ significantly between the control groups. Further evidence that sign production is not necessarily related to total productivity is furnished by the fact that the mean number of homosexual signs given by the three homosexual Ss who produced the most total responses was not as high as the mean number of signs for the homosexual group as a whole.

Thus, Wheeler's scheme of 20 signs seemed to be successful in differentiating overt homosexuals from neurotics and normal males, while not differentiating between the latter two groups. In addition, it was found that the overt homosexuals produced an average of 23 more responses to the Rorschach test than did the hygiene group, and 17 more responses than did the normal Ss. This finding is in agreement with Aronson's (1) suggestion that creative productivity on the Rorschach may be a homosexual variable. Since, with productivity controlled, the mean percentage of signs for the homosexual group was significantly higher than the means for the other two groups, it is concluded that the high response productivity of the homosexuals was not a vital factor in their production of more proposed signs of homosexuality.

Validity of individual signs. Although the 20 signs, as a scheme, differentiated between the homosexual and nonhomosexual groups, only 4 of the individual signs discriminated significantly between the homosexuals and the normal Ss. These individual signs are *Sign 7* ($t = 1.70$), *Sign 8* ($t = 2.17$), *Sign 10* ($t = 1.65$), and *Sign 19* ($t = 1.80$). None of these signs differentiated significantly between the two nonhomosexual groups.

The validity of *Sign 7* (Blot IV seen as contorted, monstrous, or threatening) is consistent with the clinical interpretation of Card IV as a stimulus that frequently evokes signs of castration anxiety. It is not readily apparent from the rationale offered by Wheeler why *Sign 8* (humanization of animal) should be so successful, but both Wheeler (13) and

Reitzell (10) identified it as an effective sign. *Sign 10* (derogatory attitude toward females) and *Sign 19* (preoccupation with genitalia) have an obvious relation to homosexual dynamics.

It should be emphasized that these signs are listed because they show a *significant degree of differentiation*, not because of their frequency of occurrence in the homosexual protocols. That is, the frequency of occurrence of a sign was not always indicative of its ability to differentiate.

Most of the remaining signs proved useful, to some degree, although many of them did not occur frequently enough to be validated. Only one sign can be seriously questioned. *Sign 1* (mask or face on Card I) was given 11 times by the homosexual group and 11 times by the normal group. The frequency of occurrence of this sign, in normal as well as homosexual Ss' protocols, suggests that intellectualization of Card I into a face may not be a valid discriminator, and the usefulness of this sign awaits further experimental evidence.

Validity of individual cards. Four of the 10 Rorschach cards differentiated between the homosexual group and the normal group, in the predicted direction, beyond the .05 level of significance. These are Cards II ($t = 1.97$), IV ($t = 2.05$), V ($t = 1.80$), and VII ($t = 1.70$). None of these cards differentiated significantly between the two nonhomosexual groups. The results obtained with Cards IV and V were due largely to the success of *Signs 7 and 8*, which accompany these cards. The present findings with Card IV support the clinician's faith in this card as securing responses indicative of confusion and unresolved conflicts in regard to relations with males in general and "father figures" in particular. Card II elicited genital, anal, and religious responses, while Card VII evoked responses derogatory to women, and descriptions of animals or humans as being "back to back."

Signs suggested for further confirmation. It was noted that the homosexual Ss gave many responses involving the *rear view* of human beings or animals. For example, responses like "figure seated with its back to you," "back of two legs," or "Negro can-can," were given quite frequently by the homosexuals. Com-

parison of the mean frequency of rear-view responses given by the homosexual group and the normal control group resulted in the highly significant t of 2.53. This *rear-view* sign is probably a more defended form of Wheeler's *anal detail* sign. The latter sign did not occur very frequently in the present study. At any rate, the finding of a significant difference between the homosexual and normal Ss suggests that the rear-view sign may prove to be a valid index of homosexuality. This new sign is offered for further investigation and confirmation.

Discussion of Rorschach findings. Although content analysis seemed useful here, certain deficiencies of this type of analysis became apparent to the investigators. Much of the distinctiveness of the homosexual protocols could not be measured by the sign method. The homosexual reactions to the blots, as reflected by personal remarks and the use of particular words recorded in the data, seemed to display a great sensitivity of the S toward the blots and a need to relate each blot to himself. Remarks such as "I don't like that one," or "he looks mean," were frequent for the homosexual Ss, while the control Ss gave responses that were more detached. This sensitivity and greater ego-involvement toward the Rorschach test is, in part, reflected by the greater productivity of the homosexuals. Also, the homosexuals displayed a greater "shock" reaction to color on the Rorschach than did the control Ss. These and other nuances, which prove most useful to the clinician, are overlooked by systems, such as the present one, of sign scoring.

As a whole, however, the signs were successful in differentiating between the homosexual group and the control groups, but the dangers of applying the scheme blindly to an individual protocol were obvious. One member of the normal group gave seven signs, while another gave six, both being above the mean number given by the homosexuals. Examination of the complete case records of these two control Ss revealed no evidence of homosexual tendencies. On the other hand, one homosexual S gave only two sign responses, which is below the mean for the normal Ss. Thus, there is considerable danger of making a grave mistake in diagnosis if the signs are applied as criteria without supplementary information.

TABLE 4
VALIDITY OF TAT SIGNS OF HOMOSEXUALITY

Cards 4, 6, 7, 10, 18	Homosexual Group (<i>N</i> = 15)	Normal Group (<i>N</i> = 20)	<i>F</i>	<i>t</i>
Mean number of signs	2.73	.85		3.71**
Variance	3.21	.87	3.70**	
Cards 4, 6, 7, 18	Homosexual Group (<i>N</i> = 15)	Neurotic Group (<i>N</i> = 20)	<i>F</i>	<i>t</i>
Mean number of signs	1.67	.63		4.08**
Variance	.95	.36	2.64*	
Cards 1, 3, 4, 6, 10, 11, 13, 14, 16, 18	Homosexual Group (<i>N</i> = 5)	Normal Group (<i>N</i> = 20)	<i>F</i>	<i>t</i>
Mean number of signs	3.80	2.40		1.87*
Variance	2.20	2.25	1.02	

* Significant at the .05 level for a one-tailed test.

** Significant at the .01 level for a one-tailed test.

Thematic Apperception Test

The lack of consistency in administering TAT cards to the various Ss has been mentioned. However, several comparisons were made possible by forming subsamples from the various groups, the size of these subsamples and the number of cards employed varying in each case in order to make the most efficient use of the available data.

A comparison involving the greatest possible number of Ss was made between 15 homosexual Ss and 20 normal Ss on the basis of responses given to cards 4, 6, 7, 10, and 18. With this equitable five-card comparison, the homosexual group mean of 2.7 signs was significantly higher, as shown in Table 4, than the normal control group mean of .9 signs.

The best comparison possible employing the neurotic group was made on the basis of responses given to Cards 4, 6, 7, and 18 by the same 15 homosexual Ss and the 20 neurotic Ss. As shown in Table 4, for these four cards, the homosexual group mean of 1.7 was sig-

nificantly greater than the neurotic group mean of .6 signs.

These two comparisons, although based on responses made to a small number of cards, do employ cards that were selected because they seemed most likely to elicit material relevant to homosexuality. The special relevance of Card 10 is evidenced by the considerable reduction in the homosexual group mean in the second comparison. In any event, these two comparisons clearly confirm Hypothesis II.

One additional comparison was made in an effort to take into consideration responses given to a greater number of cards. Five of the homosexual Ss had composed stories for Cards 1, 3, 4, 6, 10, 11, 13, 14, 16, and 18, and these responses were compared with the stories given by the 20 normal control Ss in response to these same 10 cards. A *t* of 1.87 resulted, which is statistically significant but considerably lower than the other *t* values presented in Table 4. On closer examination, two factors seemed to account for the reduced *t*. The inclusion of Card 13 enabled the normal Ss to give numerous sexual references, while the inclusion of Card 3 led to increased misrecognition of sex in this normal group, a distortive factor which will be discussed later. These sexual references and misrecognitions, which were encouraged by the nature of the stimulus pictures, although also increasing the number of signs produced by the homosexual Ss, decreased the significance of the proportional difference in signs between the two groups.

Validity of individual signs. Since comparisons were based on only a few cards, and since the TAT signs did not occur as frequently as did those in the Rorschach protocols, it could not be expected that many of the TAT signs would individually be capable of differentiating between the groups. However, two signs (*Signs* 5 and 10) did discriminate significantly. The results with these two signs are based on the five-card comparison (Cards 4, 6, 7, 10, and 18) between the 15 Ss in the homosexual group and the 20 Ss in the normal control group.

The most discriminatory sign was *Sign* 10 (sexual, genital references). The homosexuals gave more sexual references (*t* = 3.46) than did the normal Ss. Here, a particularly note-

worthy finding was that among the sexual signs scored for the homosexual Ss on these five cards, there were seven explicit homosexual stories. Moreover, two additional stories of homosexual relations were told in response to other cards. These nine stories were told by six of the homosexual Ss. No such stories were found in any of the data available on either of the two nonhomosexual groups. These homosexual stories revealed a wealth of information about the dynamics of the storyteller, but their diagnostic value is undoubtedly of limited generality, since such stories are likely to be told only by admitted overt homosexuals who are aware that the examiner knows the nature of their problem. However, a *t* test between the number of sexual responses given by the two groups, omitting the seven stories of overt homosexual activities, still resulted in a significant *t* of 3.17.

The other discriminating sign was *Sign 5* (a high degree of feminine identification). The *t* in this case was 1.71, significant at the .05 level for a one-tailed test. The success of this sign is particularly meaningful for two reasons. Since identification is often difficult to distinguish from an objective attempt by the *S* to describe the pictured character fully and realistically, identification was only scored when feminine identification was quite pronounced and the storyteller showed an unusual ability to create a picture of feminine yearnings and feelings. Furthermore, of the five cards used for comparison, the three that have female characters in the situation also contain equally prominent male characters, so that sexual identification was entirely a function of the *S*. This is in contrast to a card like Card 2, in which a woman is centered in the picture, providing an obvious choice for identification. At any rate, the present finding of a greater need and ability of the homosexual Ss to identify with females is in agreement with psychological theory on homosexuality.

Sign 7 (misrecognition of sex) presents special difficulties, notably in regard to its association with Card 3. According to the TAT manual (9), the character portrayed in Card 3 is labeled as a boy; thus, any reference to this character as a female was scored as a misrecognition of sex. However, the possible ambiguity of the stimulus figure in this

picture was sharply pointed out by the fact that 9 of the 20 Ss in the normal control group described the character as a girl or woman. In addition, 10 of the 20 neurotic Ss described it as a woman, while 8 of the 11 homosexuals who responded to this card called it a woman. The incorrectness of scoring this as a sign of homosexuality is clearly illustrated by a comparison of the three proportions, which results in no significant differences. These findings suggest quite definitely that the figure in Card 3 may best be considered an ambiguous figure. Few misrecognitions were made by any of the Ss in response to other cards.

Therefore, the TAT signs, as a scheme, seemed to be characteristic of homosexuality, although only two of them individually discriminated significantly. These two, sexual references and feminine identifications, were found especially useful when applied to Cards 4, 6, 7, 10, and 18, which, unlike some of the others, do not confront the *S* with a stimulus situation necessarily requiring a sexual (genital) plot or identification with a female. The validity of the other signs was found to be either limited or unsatisfactory. However, the distinctive quality of the homosexual protocols suggested that other differentiating signs were there to be found.

Validity of individual cards. Of the five cards that, because of their theoretical relevance, were administered uniformly to the majority of the homosexuals, four attracted significantly ($p = .05$) more homosexual signs from the homosexual group than from the normal control group. The best card proved to be Card 10 ($t = 2.49$), which elicited sexual references, feminine identifications, and stories of marital discord and pretense from many of the homosexual Ss, while most of the normal Ss composed "happy marriage" stories, and few of them identified with the woman in the picture. Card 7 was also successful in differentiating between the groups ($t = 2.27$), although much of this success came from its ability to elicit stories of overt homosexual relations from the homosexual Ss. Cards 4 and 6 also differentiated significantly ($t = 1.77$ and 1.90 respectively), again mainly because of the sexual references, feminine identifications, and derogatory attitudes toward women and marriage. Card 18 resulted in differences in the predicted direction, but did not dis-

criminate significantly ($t = 1.32$). It is possible, of course, that some cards in the female series (e.g., 18GF) might be worth experimenting with in future studies of male homosexuality.

Signs suggested for further confirmation. Five new signs were suggested for further validation, three of them significantly differentiating homosexuals from normals, and the other two not occurring as frequently but worthy of mention for their possible diagnostic value.

The first sign suggested for further research applies to stories containing a *strong unresolved attachment to a father or father figure*. Scored in stories composed for Card 8, this sign resulted in the highly significant t of 2.54. It was only scored when the attachment to the older man was the focal point, not merely incidental to other facets of the story. Remarks such as, "He couldn't bear to leave his father," and "The older man and the younger man had a strong relationship, because the old man took a great deal of interest in him," were frequent in the homosexual protocols. This suggested sign has both a theoretical and an empirical basis, the observation of homosexual relationships between a young man and an old man being theoretically referable to the unresolved Oedipal conflict.

Another sign, associated with Card 6, applied to stories involving a *strong unresolved attachment to the mother*. Comparison of the homosexuals and the normal Ss resulted in a significant t of 1.96. Here again, the sign was only scored when the mother-son relationship was described as being "a silver cord attachment," or from statements like, "They had always been too close, if anything." Moreover, one explicit mother-son incest story was produced by a homosexual S. Although the normal Ss told of mother-son relationships in response to Card 6, the focus was usually a financial problem, a "guilt" theme the boy's leaving his mother to marry, or a sense of duty or feeling of unworthiness in the son's relationship to the mother. Whenever a story implied an unnatural love toward the mother, it was found to be produced by a homosexual S. The success of this sign is understandable, theoretically, in terms of an unsatisfactory resolution of the Oedipal love for the mother.

Although the homosexuals did not often produce stories in which a clearly derogatory attitude toward women was revealed, use of

certain *sexual derogatory terms in relation to women* gave specific clues to their feelings toward female characters in their stories. Words like "tramp," "babe," "temptress," "vamp," "floozy," "wench," and so forth were far more frequent in the homosexual protocols than in the protocols of the normal Ss. A comparison of the mean number of such terms used by the homosexuals and the normal Ss resulted in a significant t of 2.31. It is noteworthy that in many of the stories employing these words, the homosexuals described another female character in a favorable light, so that the derogatory attitude was not evident in the story as a whole, but was betrayed only by the occasional word.

Two other possible signs, which did not occur with sufficient frequency to be validated but were notably absent from the normal Ss' stories, seem worthy of further investigation. The homosexuals gave several stories in which a *heterosexual relationship was followed by tragedy*, usually the violent and sudden death of one member of the couple. Also, in response to Card 18, several homosexuals told stories that *describe the plight of the young man allegorically or symbolically*, defining the situation as a "symbol of revolt against society" or the "torture of the damned." This allegorical description, which is in contrast to the realistic stories told by the normal Ss, is possibly an intellectualization of the Ss' homosexual conflict.

Discussion of TAT findings. As in the Rorschach, the examiner could often feel an immeasurable distinctiveness in the homosexual TAT data. This feeling was unbiased since the stories were first scored blind, with the scorer able only to guess to which group a protocol belonged. There was, again, in the homosexual protocols, a deeper sensitivity than was displayed by most of the normal Ss, a fuller projection into the story, including the ability to empathize with the characters. This sensitivity was accompanied by elaborate descriptions of the emotional relationships between various characters in the story. Also, there was often a sense of fatalism, of tragic events to come which could not be evaded by the characters. These qualities, which are difficult to describe objectively and to quantify, are meaningful to the clinician but are difficult to record by the present sign scheme.

Again, the uses and limitations of sign

analysis are evident. The success of the scheme in differentiating between groups is offset by the dangers involved in judging the individual case by the number of homosexual signs in the protocol. One normal *S* gave seven TAT signs of homosexuality, considerably above the mean number for the homosexuals, yet his complete file was found to be free of evidence of homosexuality. As with the Rorschach, TAT signs should be used only in conjunction with supplementary evidence gained from the case history and other psychodiagnostic procedures. They definitely cannot be expected to carry the burden of complete diagnosis of the individual case.

Joint Use of the Rorschach and Thematic Apperception Test

Hypothesis III predicted a significant positive correlation within groups between the number of Rorschach and TAT signs of homosexuality given by each *S*. It would be expected that if the Rorschach and TAT signs were each capable of detecting this abnormality, they would agree in their measurement of the relative prominence of the variable in each *S*.

Product-moment correlations (r) were computed within each group to determine the degree of association between results of the two tests. With the neurotic group, the resulting r of $-.09$ does not permit rejecting the null hypothesis of no association; and with the normal control group, the correlation of $.05$ fell far short of significance.

Since the degree of correlation of the two tests for the homosexual group is of crucial interest in this study, the TAT results were ordered into two schemes for correlation with the Rorschach signs. First, the complete sign totals on both tests for all 20 *Ss* were correlated, although the TAT signs of many of these *Ss* were based on a varying number of cards, ranging from one to fifteen. It could be expected that the disparity in the completeness of the TAT data on each *S* might prove detrimental to obtaining the predicted positive relationship. However, in spite of this inequity, a significant r of $.52$ resulted from this comparison.

Then a more equitable correlation was attempted within the homosexual group. The TAT sign totals of the 15 *Ss* who had produced stories for the same five cards were correlated with the sign totals of these *Ss* on the Ror-

TABLE 5
CORRELATION BETWEEN HOMOSEXUAL SIGNS ON THE RORSCHACH AND TAT

Group	N	r
Neurotic	20	$-.09$
Normal	20	$+.05$
Homosexual	20	$+.52^{**}$
Homosexual	15	$+.51^*$

* Significant at the .05 level for a one-tailed test.

** Significant at the .01 level for a one-tailed test.

schach test, resulting in a coefficient of $.51$. Thus, both correlations for the homosexual group proved positive and significant.

Although Hypothesis III was confirmed only in the case of the homosexual *Ss* and unsupported by the records of the control *Ss*, these results, shown in Table 5, provide a clear insight into the success and limitations of this investigation. Apparently the two tests did what they were expected to do. They discriminated a group of homosexuals from two groups of nonhomosexual males. In addition, they measured the relative prominence of the homosexual characteristic in a similar fashion for the homosexual *Ss*. However, to expect the tests to correlate significantly in detecting signs of latent homosexuality in nonhomosexual *Ss* is not only drawing too fine a line for differentiation, but also entails the assumption that each sign is a relatively foolproof detector of the homosexual variable. The results of this experiment certainly demonstrate the fallacy of considering any sign a definite detector, as well as point out the varying validity of the signs. In the case of the homosexuals, the possible lack of validity of a given sign is usually offset by the large total number of signs, which is somewhat more valid and less susceptible to error.

DISCUSSION AND CONCLUSIONS

While these two schemes appear useful in characterizing homosexuality, there are, nevertheless, certain limitations that must be kept in mind, both in accepting the validity of these indices and in deriving generalizations from them. The most glaring and serious limitation is that these results do not permit the conclusion that a large number of these signs in the Rorschach or TAT is necessarily an indication of homosexuality. The case of

the normal *S* with several signs in his protocol, as well as that of the homosexual *S* with very few signs, definitely occurred. This possibility is to be expected, since several of the signs, in either scheme, cannot be considered unusual responses, and some could be interpreted in terms of other personality variables. For example, in one record "anal" imagery may be symptomatic of homosexuality, while in another the same imagery may be indicative of an obsessional theme. Even feminine identification, a useful sign in the TAT and the rationale for a Rorschach sign, when based on identification with the mother, is not infrequent in American society and is reflected in the TAT stories of many heterosexual males (12).

It is possible that some signs proved useful in this study only because the homosexual *Ss* were highly intelligent. Perhaps, for example, a group of homosexuals of different intellectual capacity would not humanize animals on the Rorschach test. Since the present design does not permit the rejection of such possibilities, it seems wise to restrict the import of these findings to men of college age, of high average socioeconomic background, and of high intelligence. However, as the homosexual variable was still the most basic, it seems plausible that these results and considerations may hold true for homosexuality in general.

The use of a group of overt homosexuals permitted a critical test of the validity of the proposed signs. But the use of this group of *Ss* also presents a difficult problem in the interpretation of the results. Since these *Ss* all came to be tested because they wanted to talk about their problem and all were aware that the examiner knew their reason for coming, it is to be expected that their test responses would be frank and unsuppressed. Therefore, much of the knowledge gained from studying this homosexual group may be of little avail in investigating the protocols of other individuals who have no desire to be frank and who may feel themselves at odds with the psychologist. In other words, it is difficult to say how much of what has been learned here will prove useful in detecting a latent homosexual conflict or in identifying a homosexual who is deliberately trying to prevent discovery of his problem.

In reviewing both the historical background

of the problem and the results of this study, it has become obvious that Rorschach research on homosexuality has been more complete and effective than what has been done with the TAT. The Wheeler scheme, in comparison with the TAT list of signs, is more complete, more specific, and seems to differentiate better between the groups.

Limitations of the TAT sign approach were made clearly evident by several examples in the protocols. A bizarre story told by a homosexual in response to Card 13 about a boy who shares his girl with his roommate is scored as a sexual reference along with other more conventional sexual stories told by control *Ss*. Full use and interpretation of such meaningful material cannot be handled by a system of sign scoring, especially one so crude as that used in this experiment. However, the distinctive quality of the homosexual TAT responses suggests that additional signs of a definite nature can be postulated and may prove useful. This study has suggested five such signs. If, in a future study, a greater number of TAT cards is administered to a homosexual group, more such relevant material would probably be uncovered.

The success of various signs in both projective tests contains several implications for theory on the topic of homosexuality. A preoccupation and lack of satisfaction with sex is brought out by the many genital responses given on the Rorschach by the homosexuals as well as by their inclusion of sexual themes and derogatory sexual terms on the TAT in response to stimuli which are not specifically structured to elicit such material. The derogatory attitude of these homosexuals toward women in general is evidenced both on the TAT and the Rorschach. Paradoxically, this depreciation of women seems to be often accompanied by a feeling of identity with them, manifested not only on the TAT, but on the Rorschach as well.

Oedipal implications are involved in the stories of unnatural love and overclose ties to the mother. An interesting note here is that in most of the stories told by homosexuals of this abnormal closeness to the mother, there was usually a deprecating tone in reference to the attachment, as if the *S* were admitting this tie grudgingly and as if he were trying desperately to break away from it. This is man-

ifested in references such as "the umbilical cord was never broken." In contrast to the depreciation of the overclose mother tie, bonds to another man were usually idealized. The older man was "very kind," or "really interested in him," and there was contentment for both parties in the male relationship. Thus Oedipal implications concerning the homosexual S's identification with the mother are evidenced by these stories of "love for an older man." The importance of parental figures in shaping later attitudes toward and identification with the sexes was brought out again and again.

The confusion and conflict in the homosexual role is indicated by the success of several of the signs. The frequent religious references in response to the Rorschach inkblots suggest the heavy burden of guilt and self-questioning that the homosexual must bear. The allegorical references on the TAT are probably also an expression of the homosexual's inner conflict. The humanization of animals on the Rorschach test may also suggest a confusion of role and sexual identity, or even of body image, as a source of anxiety.

One further point seems worthy of mention. It was observed that the responses of the homosexuals on both tests ranged from the most romantic and unrealistic, such as symbols of royalty on the Rorschach and tales of the "tragic hero" and stories of "another world," to the most lurid and earthy Rorschach responses like "bleeding rectum" or "vaginal disease" and TAT stories of incest and suicide. This approach to the projective stimuli may represent a manifestation of the homosexual conflict between high sensitivity and craving for beauty and all the disgust and disfavor which society associates with the homosexual's problem. Both extremes imply a lack of realism in thinking.

This discussion points to the wide possibilities of both a "sign" analysis and a more global clinical appraisal of projective protocols. The Rorschach research has progressed further than that with the TAT, but additional investigation and validation of indices on both tests is needed. As the list of valid signs of homosexuality increases and unsatisfactory ones are eliminated, this approach to diagnosis and understanding of the homosexual problem will become less crude, and may well have more

implications for the individual case. At present, however, this content "sign" analysis should only be used to supplement the more structural and inferential appraisal of projective protocols in conjunction with any other diagnostic material available.

SUMMARY

Rorschach and TAT protocols of a group of overt homosexuals, a group of neurotics, and a group of normal males were scored for frequency of occurrence of proposed signs of homosexuality. As hypothesized, it was found that the homosexual group gave a significantly greater mean number of the Rorschach signs than did either nonhomosexual group. Four individual signs, and four separate Rorschach cards were found to differentiate significantly between the homosexual and the normal group. One new Rorschach sign uncovered in this study was offered for further confirmation. As predicted, it was also found that the homosexuals gave a significantly greater mean number of the proposed TAT signs than did either nonhomosexual group. Two individual signs were significantly discriminatory, and five separate TAT cards were recommended for administration in future studies of homosexuality. Five new signs were suggested, along with their theoretical implications, as aids to formulation of a more discriminating scheme.

Within the homosexual group, correlation between the number of Rorschach signs and number of TAT signs produced by each S proved significant, serving as a check on the validity of both schemes and indicating the consistency of these two diverse measures of homosexuality. However, similar correlations within each of the nonhomosexual groups did not approach a level of statistical significance, indicating that the signs are probably too crude to use as instruments to detect varying degrees of latent homosexuality among nonhomosexual males.

Limitations on generalizations from the present results were discussed, and the dangers involved in expecting these signs to serve as unaided indicators of homosexuality in an individual protocol were emphasized. It was concluded that the existing sign schemes must be further refined, validated, and lengthened—a process in which this experiment is only a first step.

REFERENCES

1. ARONSON, M. L. A study of the Freudian theory of paranoia by means of the Rorschach test. *J. proj. Tech.*, 1952, **16**, 397-411.
2. BERGMANN, M. S. Homosexuality on the Rorschach test. *Bull. Menninger Clin.*, 1945, **9**, 78-83.
3. COMMITTEE ON COOPERATION WITH GOVERNMENTAL (FEDERAL) AGENCIES OF THE GROUP FOR THE ADVANCEMENT OF PSYCHIATRY. *Report on homosexuality with particular emphasis on this problem in government agencies*. Topeka: 1955, Report No. 30.
4. DUE, F. O., & WRIGHT, M. E. The use of content analysis in Rorschach interpretation: differential characteristics of male homosexuals. *Rorschach Res. Exch.*, 1945, **9**, 169-177.
5. KINSEY, A. C., POMEROY, W. B., & MARTIN, C. E. *Sexual behavior in the human male*. Philadelphia: Saunders, 1948.
6. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. New York: World Book Co., 1942.
7. LINDNER, R. M. Content analysis in Rorschach work. *Rorschach Res. Exch.*, 1946, **10**, 121-129.
8. LINDZEY, G., TEJESSY, CHARLOTTE, DAVIDS, A., & HEINEMANN, SHIRLEY. Thematic Apperception Test: a summary of empirical generalizations. Unpublished manuscript, Harvard Univer., 1952.
9. MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Univer. Press, 1943.
10. REITZELL, JEANNE M. A comparative study of hysterics, homosexuals, and alcoholics using content analysis of Rorschach responses. *Rorschach Res. Exch.*, 1949, **13**, 127-141.
11. TEJESSY, CHARLOTTE. The Thematic Apperception Test: a summary and preliminary investigation of empirical generalizations. Unpublished honor's thesis, Harvard Univer., 1952.
12. TOMKINS, S. S. *The Thematic Apperception Test*. New York: Grune & Stratton, 1947.
13. WHEELER, W. M. An analysis of Rorschach indices of male homosexuality. *Rorschach Res. Exch.*, 1949, **13**, 97-126.

Received July 7, 1955.

THE EFFECT OF PSYCHOPATHOLOGY ON VISUAL DISCRIMINATION¹

HERMAN NIEBUHR, JR.

University of Pennsylvania, Medical School

AND DAVID COHEN

VA Hospital, Coatesville, Pennsylvania

THE effect of psychopathology on perceptual processes is relatively unexplored. Inferences about the perceptual functioning of the psychologically ill have generally been drawn from performance on psychological tests. For the most part, such clinical inferences tend to be rather speculative, for behavior on a psychological test is compounded of several functions. Rorschach performance, for example, is not solely a perceptual act, but involves the complex analytical and integrative skills usually associated with thinking. Visual motor performance tests also are compounded of several psychological functions. Consequently, inferences concerning perceptual functioning on these tests are confounded by the presence of extraneous variables and require caution in their evaluation.

The present study of visual-motor processes attempts to determine the contribution of the perceptual component—specifically, visual discrimination—to performance on a well-known clinical instrument, the Bender-Gestalt Test. Bender (1) and others (3, 6), in discussing the functions measured by the Visual Motor Gestalt Test, have placed primary emphasis on the “visual-motor function” per se. Korchin and Basowitz (4) dealt with the perceptual aspects in their tachistoscopic study but again used motor reproduction as their measure of perception. Research on this instrument has in general been directed toward the construction of scoring schemes (2, 5) with little effort to delineate the locus of the distortion one finds in the reproductions of psychopathological subjects. The question may be raised whether patients perceive the designs accurately and then distort them in reproduction, or whether they actually misperceive the design and then more or less accurately reproduce their misperceptions. In the present study, the perceptual aspect of the Bender-Gestalt Test was studied under memory and

copy conditions by means of a multiple-choice test. Motor reproduction aspects were also investigated under memory and copy conditions. Comparisons were made between perceptual and motor performance under both conditions.

METHOD

Subjects. There were 40 Ss: 10 student nurses, 10 acutely schizophrenic patients, 10 chronic schizophrenic patients, and 10 patients hospitalized for neurological reasons. All patients were hospitalized in a Veterans Administration psychiatric hospital at which the student nurses were being trained. Characteristics of the sample are given in Table 1.

Although Pascal and Suttell (5) suggest that age and intelligence are not significantly related to Bender-Gestalt performance, an attempt was made to equate groups for intelligence as far as possible. No attempt was made to control age. All patients were in sufficient contact to cooperate in the experiment, and each S completed all tasks.

Materials. Test materials consisted of a standard set of Bender-Gestalt cards and a specially constructed multiple-choice test in which each S was required to select from six designs the one which was the exact reproduction of the standard stimulus. A series of alternates was constructed for each of the nine standard designs. In each series, the alternates were of varying similarity to the standard design. No attempt was made to rank the difficulty or similarity of the five wrong choices in each series. Paper cut to the size of the standard Bender-Gestalt stimulus cards was used for the motor reproduction tasks.

Procedure. Each S was asked to perform four tasks in the following sequence. There was no delay between these four steps.

I. *Memory Reproduction:* Ten-second presentation of each Bender-Gestalt design from A to 8. After each presentation, S was requested to reproduce the design from memory.

II. *Copy Reproduction:* Presentation of each Bender-Gestalt design from A to 8. The S was requested to reproduce each design by copying. There was no time limit, and the standard design was kept before S until completion.

During the reproduction tasks, each S was instructed to reproduce the standard designs as accurately as possible and was cautioned to pay attention to the size of his reproduction, the accuracy of all its parts, and its correct placement on his drawing paper so that his drawings were to be as exact a reproduction of the standard designs as he could possibly effect. The importance of accuracy was reemphasized upon presentation of each succeeding design.

¹From the Veterans Administration Hospital, Coatesville, Pennsylvania.

TABLE 1

CHARACTERISTICS OF EXPERIMENTAL AND CONTROL GROUPS

Group	Mean Age	Mean IQ	Mean Time Hospitalization	Diagnosis	N
Student Nurses	19.7	101.5*	—	—	—
Acute Schizophrenic	28.5	102.2**	6.1 mos.	Paranoid Mixed Catatonic	7 2 1
Chronic Schizophrenic	36.0	96.3**	7.7 yrs.	Paranoid Hebephrenic Catatonic	5 2 3
Neurological	59.1	Not available	9.9 yrs.	Arteriosclerosis Multiple sclerosis Chronic brain synd. Tertiary syphilis General paralysis Epilepsy, grand mal	3 1 2 2 1 1

* IQ derived from California Short-Form Test of Mental Maturity, Advanced Form.

** IQ derived from Wechsler-Bellevue Adult Intelligence Scales.

III. *Memory Multiple Choice*: Immediately following the preceding two steps and without further experience with the standard designs, *S* was requested to select each standard design from memory from a series of six alternates for each of the nine designs.

IV. *Matching Multiple Choice*: With each standard design before him, *S* was asked to select the correct design from the same series of six alternates used in Step 3 by matching each alternate with the standard design.

Scoring. For the multiple-choice tasks, a score was obtained for each *S* by tallying the number of correct selections under the memory and matching conditions. The maximum score was 9 and the minimum score was 0 for each condition. For the reproduction tasks, under both memory and copy conditions, each was rated for compliance with the experimental instructions along five general dimensions: size, general placement, angular placement, detail accuracy, and design accuracy. Ratings were made on a 3-point scale, a score of 0 indicating congruence with the original design, a score of 1 indicating minor deviation, and a score of 2 indicating major deviation. Thus, there was a minimum score of 0 and a maximum score of 10 for each design reproduction. For each memory and copy condition, a minimum of 0 and a maximum of 90 points were possible. The higher the score, the less adequate were the reproductions. Specific directions for scoring each of the five variables for each design, as well as scoring examples, were presented to the rater in a Rater Instruction Sheet. A more detailed exposition of the scoring dimensions follows:

I. *Size*. Each design was rated for its size as compared with the original. A reproduction could be larger,

smaller, or the same size as the original. A score of 1 was given when the reproduction was slightly larger or smaller than the original. A score of 2 was given when the size was less than half or greater than twice the size of the stimulus.

II. *General Placement*. Each design was rated for its placement on the copy paper in comparison to the placement of the standard design on its card. A score of 1 was given for deviations less than $\frac{3}{4}$ " along either the horizontal or vertical dimensions. A score of 2 was given for deviations greater than $\frac{3}{4}$ " along either or both dimensions.

III. *Angular Placement*. Each design was rated for its angular placement irrespective of its general placement on the copy sheet. A score of 1 was given for a rotation of part or all of the design of less than 45 degrees. A score of 2 was given for rotation over 45 degrees. Thus, severe rotations and reversals were given a score of 2.

IV. *Detail Accuracy*. Each design was rated for the accuracy with which the details of each design were drawn. This did not include the over-all or gestalt accuracy covered by the Design Accuracy score. Since there are different details for each design, the rating was qualitative. Scores of 1 and 2 were given on the basis of the degree of detail distortion. Distortions may take the form of flattening or exaggeration of curves, of drawing loops for dots, dashes for loops or dots, or ellipses for circles, or of increasing or decreasing the size of dots or loops. It is important to recognize that only the components of each design were rated.

V. *Design Accuracy*. Each design was rated for its gestalt quality. This was necessarily a difficult judgment, for it was intended to rate the adequacy of the design as a whole, reflecting both omissions and additions. Slight distortions in the shape of the design, including overlaps and fragmentations, received a score of 1. Major distortions, omissions, and additions received a score of 2. This dimension, as well as the preceding one, was described in special detail in the Rater Instruction Sheet.

Since one of the authors scored each reproduction with full knowledge of each *S*'s group status, and also scored each group by itself, it was necessary to check reliability by having another rater learn the rating system and score the reproductions in random order without knowledge either of *S*'s group membership or of the task instructions. Product-moment reliabilities of .97 and .93 were obtained for designs in the memory and copy conditions respectively. The data presented are based on the scores of the second rater.

RESULTS

Performance on multiple-choice tasks. Tables 2 and 3 present data on the discrimination accuracy between the experimental and control groups on the multiple-choice tasks. An analysis of variance (Table 3) and intergroup comparisons by means of *t* tests indicated the locus and significance of differences among the groups on the perceptual multiple-choice test under memory and matching conditions.

TABLE 2
PERFORMANCE ON MULTIPLE-CHOICE TASKS*

Group	Memory Condition		Matching Condition	
	Mean	Sigma	Mean	Sigma
Student Nurses	5.60	1.78	8.40	.70
Acute Schizophrenic	5.00	1.16	6.80	1.48
Chronic Schizophrenic	3.20	.79	5.50	1.51
Neurological	2.00	1.70	2.90	1.79

* Highest score possible, 9. Lowest score possible, 0.

TABLE 3
ANALYSIS OF VARIANCE OF MULTIPLE-CHOICE TASK
SCORES OF ALL GROUPS UNDER MEMORY
AND MATCHING CONDITIONS

Source of Variation	df	Sum of Squares	F	p
Between Groups	3	207.25	20.9	.01
Between Ss, Same Group	36	119.30		
Total Between Ss	39	326.55		
Memory-Matching Condition X Group	1	76.05	104.9	.01
Pooled Ss X Condition	3	36.85	16.9	.01
Total Within Ss	36	26.10		
Total	40	139.00		
Total	79	465.55		

Despite the small number of Ss in each group, statistical significance between groups under the memory condition is of a high order with but one exception—the difference between the student nurses and the acute schizophrenics ($p = .40$). Differences between groups under the matching condition are again all significant with but one exception—the difference between the acute and chronic schizophrenics. The student nurses showed the greatest improvement from the memory to the matching condition, with 5 out of the 10 Ss obtaining perfect scores on the matching condition. No Ss in the three patient groups equalled this performance despite the fact that cooperation seemed high in all groups.

Performance on reproduction tasks. Consistent differences among the groups in the adequacy of reproduction are apparent from Table 4. Table 5 summarizes the results of an analysis of variance. Again, the differences between groups as studied by t tests are for the most part statistically significant. The difference between the nurses and the acute schizophrenics under the memory condition approaches, but does not reach, significance

($p < .20$). This is also the case in the comparison of the acute and chronic patients ($p < .10$). Under the copy condition, all groups differed significantly except for the difference between the acute and chronic schizophrenic groups ($p < .20$). All nonsignificant differences were in the expected direction. Again, the nurses improved the most from the memory to the copy condition.

A series of t tests was made among the groups of the five component scores composing the total reproduction score. In general, it was found that differences between the groups on the reproduction score under the copy condition were greater than under the memory condition. However, while the nurses differed significantly from the chronic schizophrenic and neurological groups on almost all of the five component dimensions, there were no significant differences on any dimension between the control and the acute schizophrenic groups on the memory condition. Under the copy condition, the nurses differed significantly from the acute schizophrenics as well as from the other groups on all dimensions except the angular placement dimension. Similarly, the

TABLE 4
PERFORMANCE ON REPRODUCTION TASKS*

Group	Memory Condition		Copy Condition	
	Mean	SD	Mean	SD
Student Nurses	21.6	8.87	9.80	4.19
Acute Schizophrenic	28.0	10.60	21.8	6.86
Chronic Schizophrenic	37.4	9.16	28.0	12.15
Neurological	60.4	10.26	50.8	11.35

* Best score, 0. Worst score, 90.

TABLE 5
ANALYSIS OF VARIANCE OF REPRODUCTION TASK
SCORES OF ALL GROUPS UNDER MEMORY
AND COPY CONDITIONS

Source of Variation	df	Sum of Squares	F	P
Between Groups	3	17,216.95	30.5	.01
Between Ss, Same Group	36	6,782.00		
Total Between Ss	39	23,998.95		
Memory-Copy Condition X Group	1	1,461.25	112.2	.01
Pooled Ss X Condition	3	329.05	8.4	.01
Total Within Ss	36	469.70		
Total	40	2,260.00		
Total	79	26,258.95		

acute schizophrenic group differed significantly from the neurological group on all dimensions, while the chronic schizophrenic group differed significantly from the neurological group on three of the five dimensions. There were no significant differences between the two schizophrenic groups on any of the dimensions. Thus, four of the five component dimensions provided a fair degree of discrimination between the groups. A combination of all five dimensions, however, achieved the greatest discrimination and is thus likely to prove most useful in clinical application.

Comparison of multiple-choice and reproduction tests. It seemed useful to analyze the relationship between perceptual accuracy as measured by the multiple-choice test and motor performance as measured by the reproduction score. Within-group correlations were with one exception positive but non-significant.

DISCUSSION

The results clearly indicate that perceptual efficiency, insofar as visual discrimination is concerned, decreases with severity of psychopathology. The adequacy of motor reproduction likewise decreases with severity of psychopathology. The question of the degree and nature of the relationship between perceptual and motor efficiency remains unresolved on the basis of the present data. The positive, although nonsignificant, correlations between the two functions within each group leave the matter ambiguous at this time. However, it is to be noted that age and severity of psychopathology are highly correlated in the experimental population. Hence, a decrement in both perceptual and motor functioning as a function of age is a possibility which cannot be ruled out in spite of Pascal and Suttell's (5) results.

While the memory condition was included to impose an additional burden for the patient groups and thus to insure greater differences from the nurses, it was evident that the additional pressure of the memory condition served only to decrease those differences. The inferior performance of the patient groups under the matching condition, in which there was no time limit and the comparison design was constantly present, only serves to emphasize the marked differences in perceptual functioning. The superiority of the acute schiz-

ophrenics over the chronic schizophrenic group is consistent with Zubin's (7) finding that patients whose perceptual efficiency is less impaired have a better prognosis for hospital discharge. Thus, the evaluation of a patient's perceptual functioning as a prognostic aid appears to be a fruitful area for further investigation. What the reason is for the decrement in performance in patient groups can only be speculated upon. Whether it represents a regressive phenomenon in the sense of a return to an earlier mode of perceptual development, an impairment of the attention and concentration requisite to the accurate perception of a design, or some other process, cannot be determined from this study.

The differential levels of performance of the experimental and control groups on the reproduction tasks concur with earlier clinical and experimental studies with the Bender-Gestalt Test. The methodological modifications made in the present study seem to indicate additional useful dimensions. Not only were the detail and design accuracy dimensions, like those used in other Bender scoring systems, discriminating between the groups, but the dimensions of placement and size were also useful discriminators. The use of paper of the same size as the stimulus cards seemed to provide a more accurate frame of reference against which to evaluate performance.

A refinement of this approach is planned so as to explore further the specific nature of the perceptual disturbance involved in psychopathology.

SUMMARY

Perceptual and motor aspects of the Bender-Gestalt Test performance were compared for four groups of Ss: 10 student nurses, 10 acute schizophrenic patients, 10 chronic schizophrenic patients, and 10 patients hospitalized for neurological reasons. All patients were hospitalized in a psychiatric hospital at which the student nurses were being trained. After the Ss had obtained equal experiences with the Bender designs by reproducing them under memory and copy conditions, a perceptual multiple-choice test was administered under memory and matching conditions. Results indicate that there are marked differences in perceptual accuracy between the four groups: the control population was most accurate, with

the acute schizophrenic, chronic schizophrenic, and neurological patients less proficient in that order. Statistical significance was obtained on all but two comparisons. The results of the motor reproduction performance are similar to perceptual results.

The question of the degree and nature of the relationship between perceptual and motor efficiency remains unresolved in this study. Since age and severity of psychopathology are highly correlated in the experimental populations, one cannot rule out the possibility that decreasing perceptual and motor efficiency is a function of age in spite of Pascal and Suttell's (5) results.

REFERENCES

1. BENDER, LAURETTA. A visual motor test and its clinical use. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1938, No. 3.
2. BILLINGSLEA, F. Y. The Bender Gestalt: an objective scoring method and validating data. *J. clin. Psychol.*, 1945, 1, 134-140.
3. HALPERN, FLORENCE. The Bender Visual Motor Gestalt Test. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques*. New York: Prentice-Hall, 1951. Pp. 324-340.
4. KORCHIN, S., & BASOWITZ, H. The tachistoscopic Bender-Gestalt test. Paper read at Amer. Psychol. Ass., New York, Sept., 1954.
5. PASCAL, G. R., & TUTTELL, B. J. *The Bender-Gestalt Test: quantification and validity for adults*. New York: Grune & Stratton, 1951.
6. WOLTMANN, A. The Bender Visual-Motor Gestalt Test. In L. E. Abt & L. Bellak (Eds.), *Projective psychology*. New York: Knopf, 1950. Pp. 322-356.
7. ZUBIN, J., WINDLE, C., & HAMWI, V. Retrospective evaluation of psychological tests as prognostic instruments in mental disorders. *J. Pers.*, 1953, 21, 342-355.

Received August 12, 1955.

THE ROTATION OF DRAWINGS BY BRAIN-DAMAGED PATIENTS¹

AUBREY J. YATES

Institute of Psychiatry, Maudsley Hospital, London

THE block design rotation effect was first observed in psychiatric patients on the Goldstein Cube test, which requires the S to reproduce patterns with one-inch colored cubes. Some patients, while completing the designs correctly, left the completed pattern in a rotated position, apparently without awareness of the rotation (6). Such rotations frequently reached 45° (Fig. 1) but rarely exceeded it.

In previous work carried out by Shapiro (6, 7, 8), an attempt was made to determine the factors which influence the appearance of the rotation effect in both normal and abnormal Ss and to discover the factors which facilitate its appearance in exaggerated form in certain Ss.

The principal findings to date are of two kinds. First, brain-damaged psychiatric patients rotate significantly more than non-brain-damaged psychiatric patients (7); and second, the rotation effect appears under specifiable conditions. For example, when the line of symmetry of a design (defined as the line which divides the design into mirrored halves) is at an angle to the vertical axis of the total visual field, the tendency to rotation is increased. When the line of symmetry is parallel to the vertical axis of the total visual field, then the tendency to rotation is decreased. Also, when the design is in a diamond orientation, the tendency to rotation is increased; but when the design is in a square orientation, the tendency to rotation is decreased. This orientation is referred to as the figure shape. Finally, when the ground (defined as the 6" by 6" white card on which the design is placed) is in a diamond orientation, the tendency to rotation is increased; when the ground is in a square orientation, the tendency to rotation is decreased.

¹ The writer expresses his appreciation to Dr. M. B. Shapiro for his helpful criticisms in the design of the experiment. Originally part of a thesis submitted to the University of London in partial fulfillment of the requirements for the degree of doctor of philosophy, this work was made possible by a grant from the Research Fund, made available from the endowment by the Board of Governors of the Bethlem Royal Hospital and the Maudsley Hospital.

Thus, brain-damaged psychiatric patients tend to produce the rotation effect and to do so under identifiable conditions. Shapiro (8) explained these results in terms of an exaggeration of negative induction effects in brain-damaged patients. The concept of negative induction used is that developed by Pavlov (5), i.e., the notion that the development of excitatory effects in any part of the nervous system is followed or accompanied by inhibitory effects in other parts. The inhibitory aspects of negative induction are assumed to be exaggerated in brain-damaged patients. The justification for this generalization is found mainly in the work of Bender (1) and his associates.

From this reasoning, two predictions were made: First, brain-damaged patients would rotate significantly more than non-brain-damaged patients in an analogous situation. They were therefore required to draw the designs instead of making them with blocks. Such a procedure is much shorter and is considerably easier for the S; it is also less expensive to record than previous methods, which involved the use of a camera. Second, the factors influencing rotation in the blocks test would be the same in the drawing version, i.e., a diagonal line of symmetry would induce more rotation than a vertical, a diamond figure more than a square, and a diamond ground more than a square. The experiment reported was designed to test these predictions.

METHOD

Normal control subjects (N = 25). The nonpsychiatric control Ss consisted of a group of female student nurses. Their mean age was 20.00 (SD, 0.58; range, 19-23), and mean equivalent Wechsler² full-scale IQ, 116.24 (SD, 10.58; range, 92-142).

Non-brain-damaged psychiatric patients (N = 25). The psychiatric group included patients at various stages of illness and of various degrees of acuteness. In terms of psychiatric diagnosis, it included six cases of

² The equivalent Wechsler full-scale IQ was calculated by summing the WS for Vocabulary, Similarities, and Block Design and then prorating. This combination of subtests has been shown to correlate +0.937 with full-scale IQ in normal Ss (4) and +0.900 in a random sample of 100 psychiatric patients (9).

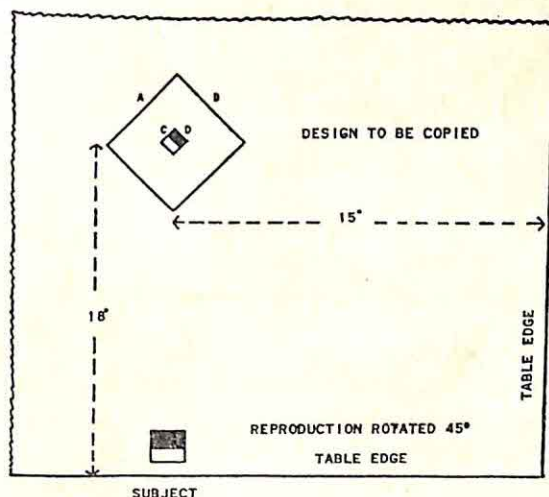


FIG. 1. EXAMPLE OF A CARD FROM THE BLOCK DESIGN ROTATION TEST WITH THE REPRODUCTION ROTATED 45°

schizophrenia, seven of depression, two of obsessional neurosis, six of anxiety neurosis, three of conversion hysteria, and one of mixed neurosis. The mean age was 28.96 (*SD*, 6.25; range, 17-45) and the mean equivalent IQ 118.08 (*SD*, 16.89; range, 76-144). There were 15 males and 10 females.

Brain-damaged patients (*N* = 20). The brain-damaged group included 12 patients with tumors, three with posttraumatic syndrome, one with aqueduct stenosis, and one with cerebral atrophy. The remaining three cases were tested after temporal lobectomy for psychomotor epilepsy. The mean age of this group was 42.35 (*SD*, 7.98; range, 26-52), and the mean equivalent IQ 104.15 (*SD*, 16.61; range, 76-139). There were 14 males and six females.

Tests. The tests employed were three subtests of the Wechsler Intelligence Scale, Form I (Vocabulary, Similarities, and Block Design), and the Drawing Version of the Block Design Rotation Test. For the new version of the test, Shapiro's (7) 40 designs were used. Instead of copying each design with four Kohs blocks, however, the *S* was required to draw it in an exercise book provided for this purpose and placed in a fixed position relative to the design. Several methods of measuring rotation were tried and rejected for various reasons. Eventually, it was decided to measure the angle of the first side drawn by the *S*.⁴

RESULTS

Association of the rotation phenomenon with brain damage. Analysis of variance and covariance carried out on the three groups for rotation and intelligence (Table 1) showed that the brain-damaged group rotated significantly more than the other two groups (at the .03

⁴ Full details of the method of administering and scoring the drawing version of the test are contained in Yates (9).

TABLE 1

ANALYSIS OF VARIANCE AND COVARIANCE
(Block design version contrasted with drawing version)

Blocks Version					
	<i>N</i>	Rotation (<i>Y</i>)		Intelligence (<i>X</i>)	
		<i>M</i>	σ	<i>M</i>	σ
Brain damaged	38	8.02	4.83	104.11	16.80
Functional	19	2.86	1.60	108.37	12.90
Control	20	4.42	3.05	113.85	13.50
Analysis of variance		$F = 13.05$ $p = <.01$		$F = 2.70$ $p = \text{NS}$	
Analysis of covariance		$F_{Y \cdot X} = 11.86$ $p = <.01$			

Drawing Version

Brain damaged	20	6.85	5.16	104.15	16.61
Functional	25	4.10	2.64	118.08	16.89
Control	25	3.31	1.04	116.24	10.58
Analysis of variance		$F = 7.17$ $p = <.01$		$F = 5.64$ $p = .03$	
Analysis of covariance		$F_{Y \cdot X} = 3.38$ $p = .03$			

level of confidence, using a one-tailed test) when the effect of differences in level of intelligence⁴ between the groups was held constant. These results may be compared with those obtained for three very similar groups using the block design version of the test (Table 1). The rotation effect with Kohs blocks discriminated at a much higher level of confidence (when intelligence was held constant) than was the case with the drawing version. It is clear, therefore, that while the results support the hypothesis that the rotation effect is associated with brain damage, the drawing version is a relatively insensitive measure of it compared with the blocks. In its present form, the overlap between groups in the drawing version is too large for the test to have much clinical usefulness.

There was no relationship between rotation and age or sex in any of the groups.

Factors influencing the appearance of the rotation effect. The principal results of an

⁴ The relationship between rotation scores and intelligence level is complex and will be the subject of a separate paper, in which a large number of groups is examined.

TABLE 2

ANALYSIS OF VARIANCE OF ROTATION— F RATIOS AND SIGNIFICANCE LEVELS

Variable	Group					
	Control		Functional		Brain Damaged	
	F	p	F	p	F	p
Figure (F)	434.201	<.01	114.191	<.01	131.145	<.01
Ground (G)	1.954	NS	7.098	<.01	57.195	<.01
Persons (P)	5.030	<.01	15.312	<.01	17.946	<.01
$F \times P$	9.482	<.01	11.629	<.01	14.280	<.01
$G \times P$	—	NS	2.447	.01	10.669	<.01
$F \times G$	—	NS	8.851	<.01	1.400	NS
$P \times G \times F$	—	NS	2.691	.01	3.560	.01

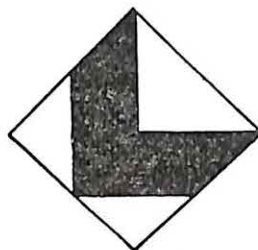
analysis of variance of the factors thought to influence rotation for each of the three groups are shown in Table 2. F ratios are not shown where the mean square for the variable was less than the residual mean square.

Contrary to prediction, there was no evidence that the line of symmetry affected rotation. Previous analyses of the data from experiments using the blocks version of the test revealed consistently from group to group that the angle of the line of symmetry in combination with the figure shape was the most important single variable influencing the appearance of rotation. When these two factors were in agreement in terms of directional value in space, rotation was maximized if both were at an angle but minimized if both were vertical. In the drawing version, however, this significant effect of the line of symmetry on rotation was not found. While the angled line of symmetry appeared to produce slightly more rotation than the vertical line of symmetry, the difference was always small and insignificant.

The most significant factor influencing rotation in the drawing version was figure shape alone. When the figure shape is square, rotation is minimal; when the figure shape is diamond, rotation is maximal, *irrespective of the direction of the line of symmetry*. In the control and psychiatric groups, the line of symmetry played no role whatever in producing the rotation effect; in the brain-damaged group, although it just attained significance level, its effect was very small indeed when compared with the influence exerted by the figure shape.

The Ss differed significantly with respect to the *relative* influence of square and diamond

STIMULUS DESIGN



REPRODUCTION

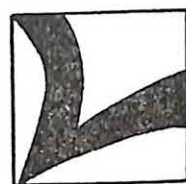


FIG. 2. EXAMPLE OF A DISTORTED OR "TWISTED" REPRODUCTION

figure shape. A much higher proportion of the total rotation score was produced by the diamond figure in some Ss than in others. This is shown by the $F \times P$ interaction in Table 2.

A diamond-shaped ground induced significantly more rotation than a square-shaped ground in the functional and brain-damaged groups, but not in the control group. The effect of ground shape was, however, relatively weak as compared with the influence of the figure shape.

Another result was quite unexpected. A number of the brain-damaged Ss first of all drew the figure shape in a rotated position. They then drew the inner content in a correct orientation (with respect to the stimulus object) but apparently without reference to the figure shape which they had just rotated. Sometimes, a form of compromise appeared to result and a distorted or "twisted" reproduction was drawn (Fig. 2). In some instances, the figure shape was distorted, but not the line of symmetry; in other instances, the reverse was the case.

DISCUSSION

It seems clear that copying designs by drawing produces essentially similar results as reproducing them by manipulating blocks. This result is consistent with other findings. Hanvik and Andersen (3), for example, reported that 59 per cent of a brain-damaged group of 44 patients produced one or more rotations of at least 30° while drawing the figures of the Bender-Gestalt test. Similarly, Hanvik (2) reported an 80 per cent incidence of abnormal EEG records in a group of 20 children showing rotations of one or more of the Bender figures.

Another finding of importance was the lack

of effect in the drawing version of the angle of the line of symmetry. Coupled with this is the observation that certain brain-damaged patients produced distorted or "twisted" copies of the designs. These two results may possibly be linked together. Nearly all *Ss* copied the designs by first of all drawing the outline, i.e., the figure shape. They then proceeded to complete the content or "inside" of the design. Thus, it was possible for an *S* to complete the outline without reference to the content and vice versa. In some brain-damaged patients, this is exactly what seemed to happen. In some instances, a kind of compromise resulted and a distorted or "twisted" design was drawn. In the blocks version, on the other hand, the situation is typically quite different. Here, the "content" and figure shape change at the same time, whichever aspect of the design is responded to. Thus, if the *S* rotates the figure shape, then the line of symmetry also rotates.

In general, the production of such "twisted" drawings may be accounted for in the same terms as the rotation effect. The brain-damaged patient may suffer from exaggerated negative induction effects in the sense of ignoring surrounding perceptual cues. In this instance, having rotated the figure shape, he then draws the content without reference to these cues. It should be pointed out that this explanation has been derived from a consideration of the results of the various experiments carried out on the rotation effect. It is therefore in the nature of a plausible hypothesis which remains to be tested.

SUMMARY

1. Rotation effects in a drawing version of a task requiring the copying of designs discriminated between brain-damaged patients

on the one hand and non-brain-damaged psychiatric patients and normal control *Ss* on the other (but to a less satisfactory degree than the blocks version of the test).

2. The relative influence of the factors affecting the appearance of the rotation effect in all *Ss* is shown to be different in the drawing version than in the blocks version. There was no influence from the line of symmetry. Figure shape exerted the main influence.

3. A tentative explanation of the disappearance of the influence of the line of symmetry and of a new finding of distorted or "twisted" reproductions is put forward.

REFERENCES

1. BENDER, M. B. *Disorders in perception*. Springfield, Ill.: Charles C Thomas, 1951.
2. HANVIK, L. J. A note on rotations in the Bender-Gestalt test as predictors of EEG abnormalities in children. *J. clin. Psychol.*, 1953, **9**, 399.
3. HANVIK, L. J., & ANDERSEN, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt test. *J. consult. Psychol.*, 1950, **14**, 197-198.
4. HILDEN, A. H., TAYLOR, J. W., & DUBOIS, P. H. Empirical evaluation of short Wechsler-Bellevue scales. *J. clin. Psychol.*, 1952, **8**, 321-331.
5. PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1946.
6. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. I. Initial experiments. *J. ment. Sci.*, 1951, **97**, 90-110.
7. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. II. Confirmatory and explanatory experiments. *J. ment. Sci.*, 1952, **98**, 605-617.
8. SHAPIRO, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, **99**, 394-409.
9. YATES, A. J. An experimental study of the block design rotation effect with special reference to brain-damage. Unpublished doctor's dissertation, Univer. of London, 1954.

Received October 3, 1955.



HOPE OF SUCCESS AND FEAR OF FAILURE AS ASPECTS OF NEED FOR ACHIEVEMENT¹

RUSSELL A. CLARK, RICHARD TEEVAN,

University of Michigan

AND HENRY N. RICCIUTI

University of Colorado School of Medicine

RESEARCH by McClelland and his associates (3) on a group TAT measure of need for achievement (*n* Achievement) yielded evidence suggesting that this measure involves two recognizable aspects of achievement motivation, *hope of success* (HS) and *fear of failure* (FF). One is an approach motive involving anticipation of reward, the other an avoidance motive involving anticipation of punishment. The main source of evidence for this distinction (summarized in [3]) consists of the repeated finding that individuals with moderate or low *n* Achievement scores appear fearful or defensively oriented whereas individuals with high scores appear hopeful. For example, those with moderate or low scores inhibit the recall of tasks failed, and are slow to recognize the tachistoscopic presentation of achievement-related words connoting failure, obstacles to achievement, etc. Subjects with high *n* Achievement scores, on the other hand, are especially quick in the recognition of achievement-related success words and show none of the defensive characteristics of the moderate group.

Clarification of these suggested relationships requires the measurement of fear of failure and hope of success, and several attempts have been made to arrive at such measures. The senior investigator first sought to arouse fear of failure and hope of success experimentally, administering the group TAT to subjects after they had been led to anticipate doing either very well or very poorly on the examination that they were about to take. Results failed to support the expectation that certain of the subcategories composing the total *n* Achievement score, singly or in combinations, would distinguish between these two experimentally different groups.

Another attempt involved a factor analysis

¹This study was supported by a grant from the Office of Naval Research under the supervision of David C. McClelland.

of the intercorrelations among these subcategories in the hope that fear of failure and hope of success might emerge clearly in such an analysis, and that categories or combinations of categories proving to be heavily saturated with the two respective factors could then be used for indices. Several such analyses were made involving somewhat different categories, combinations of categories, and criterion measures. This approach in turn was unsuccessful.

A fresh approach to the problem, a measure of these two aspects of *n* Achievement, was therefore sought that was independent of the measurement of the achievement motive. Studies of level of aspiration seemed to offer a way of isolating the cautious, defensive individual (FF) as opposed to the hopeful person (HS). A questionnaire on level of aspiration was therefore designed for the present purpose.

Because of recognized complexities in the measurement of level of aspiration, it had been given low priority in our previous approach to the problem. Our experience with such a measure seems worth reporting, however, not only because it yielded some positive results but also because it shed light on one plausible reason why some of the previous attempts had failed.

METHOD

The group TAT measure of *n* Achievement had been administered to three classes of Swarthmore freshmen under so-called neutral conditions about five weeks prior to the final examination. The level-of-aspiration questionnaire was administered to these classes immediately before the examination. It read as follows:

You are asked to cooperate in a study of students' attitudes toward examinations by answering several questions which are explained below. The information you record will be held in strict confidence, and will not be seen by your instructors, or by anyone concerned with your course grades. Please express yourself as honestly and candidly as you can since the success of the study will depend upon your giving your frank and best opinion.

Question 1. On the grade scale in the table below,

you are to indicate the two grades between which you are *practically certain* your final examination grade will fall. Make this as objective a judgment as you can, following these directions.

Lower Extreme: The lower grade you mark would answer this question: "I'm practically certain I will get at least a —."

Upper Extreme: The higher grade you mark would answer this question: "I'm practically certain I won't get a better grade than a —."

Grade Scale

E D- D D+ C- C C+ B- B
B+ A- A A+

Question 2. For each of the grades in the table above, indicate how you believe you would feel if you made that particular grade on your final examination. Do this by placing *below each grade from E to A+*, the number of the most appropriate statement below, which would describe how you would feel. Do this on the line marked "Question 2."

- +3 I would feel extremely good and very satisfied with my grade.
- +2 I would feel quite good and quite satisfied.
- +1 I would feel somewhat satisfied.
- 0 I would feel about neutral—neither good nor bad.
- 1 I would feel somewhat disappointed.
- 2 I would feel quite disappointed and dissatisfied.
- 3 I would feel extremely bad, and very much dissatisfied with my grade.

Question 3. Supposing that it were possible for you to be excused from taking the final examination you are about to take. In that case, would you be willing to settle for an A+? For each of the remaining grades from E to A+, indicate whether you *would* (Y) or *would not* (N) be willing to settle for that grade rather than take the examination, if that were possible. Do this by marking a Y or N under each grade.

Question 4. What grade are you really going to try to make on the examination?

Thank you very much for your cooperation. When you have finished, please put this sheet in the envelope and seal the envelope.

In the questionnaire, Question 1 calls for a realistic estimate of the limits within which the students' grades will fall. Question 2 calls for the *affect* that would be experienced upon receiving any particular grade. Question 3, which is somewhat of an innovation in level-of-aspiration studies, asks for the lowest grade that the student would "settle for" in lieu of taking the final exam. The fourth question is one frequently asked in level-of-aspiration studies. It asks for the grade that the individual is "really going to try to make."

How can these items be used to distinguish FF from HS? Several possibilities were entertained. First, we expected that an FF student might settle for a grade that was farther below his upper realistic limit than would an HS person (Index I). Such a defensive, fearful person should accept a lower grade in relation to his realistic expectations because doing so minimizes possibility of failure. An HS student, however, is

optimistic and should be unwilling to settle for a grade very far removed from his realistic upper limit. An equally meaningful index would take into consideration the range of expected scores. A second index therefore, was obtained by dividing Index I by the range. Still another index follows from the consideration that, other things being equal, the affect associated with the grade "settled for" should be more positive for the HS person than for the FF person. The HS person should not settle for a grade unless he is fairly well satisfied with it, while an FF individual should be willing to compromise in this respect. A fourth index is related to the third, and attempts to correct for differences in realistic expectations. The difference between the affect associated with the upper limit grade and the grade "settled for" should be less for the HS group than for the FF group. For example, an HS person who realistically expects to get a poor grade might settle for a grade close to this level without being highly pleased with it, provided that he anticipates that it would give almost as much satisfaction as the upper realistic limit. This formulation permits HS subjects to have low or negative absolute levels of affect, but requires that these levels be close to the level of affect associated with the upper limit.

In treating responses to Question 1, adjacent grades were considered to be one unit apart. The distribution of scores for Index I (Upper Limit minus Settle For) was grouped into seven class intervals, with a value of seven being assigned to the smallest difference. Since the prediction is that for the HS person the quantity (Upper Limit minus Settle For) will be small, the data were coded so that HS subjects would receive the highest scores. The results for the other three indices were treated in parallel fashion.

RESULTS

Table 1 presents intercorrelations showing the internal consistency of these indices. It shows that the correlations are all positive and highly significant. For establishing further relations with the HS-FF continuum all four categories were summed in order to obtain a more stable though crude measure. This distribution was divided into thirds with the

TABLE 1
RELATIONSHIP AMONG FOUR INDICES OF HOPE OF SUCCESS AND FEAR OF FAILURE*
(Each index is correlated with the sum of the other three)

Index	Correlation with Remaining Three Indices
Upper Limit minus Settle For	.85
Upper Limit minus Settle For Range	.82
Affect over Settle For	.58
Affect over Upper Limit minus Affect Settle For	.83

* For $N = 45$ a correlation of .38 is significant at the .01 level of confidence.

TABLE 2

RELATIONSHIP BETWEEN THE HOPE OF SUCCESS-FEAR
OF FAILURE CONTINUUM AND QUESTIONNAIRE
RESPONSES

(χ^2 test of independence)

Questionnaire Items	χ^2*	p
Settle For	27.45	< .01
Level of Aspiration	9.12	.06
Upper Limit	1.57	—
Lower Limit	8.20	< .10

* χ^2 was calculated using the method of maximum likelihood.

TABLE 3

TOTAL N ACHIEVEMENT SCORES FOR THE HOPE OF
SUCCESS, MIDDLE, AND FEAR OF FAILURE
GROUPS

(F_1 = between groups/within groups; F_2 = Middle
and combined HS, FF group/within groups)

Group	N	Mean n Achievement Score
HS	14	11.57
Middle	16	18.00
FF	15	10.73
	$F_1 = 4.26$	$F_2 = 8.44$
	$p < .05$	$p < .01$

high third showing hope of success in clearest form and the low third showing the most consistent fear of failure. The middle third might be characterized either in terms of conflict or of realistic orientation, alternatives that we are not at present in a position to distinguish.

Table 2 presents the relationship between this three-way categorization and four separate questionnaire responses: Upper Limit, Lower Limit, Settle For, and Level of Aspiration. We had made no prior predictions concerning these relationships, but, other things being equal, one would expect that the HS subjects would have higher absolute levels of Settle For and Level of Aspiration. In our prediction we had used Settle For, but the contention was that it would be closer to the Upper Limit for the HS students and not necessarily at a higher absolute value. In any event both the Settle For and Level of Aspiration turn out to be higher for the HS than for the FF subjects.

We had no expectations about the relationship with Upper Limit and Lower Limit. The HS group and the Middle group have about the same distribution at upper and lower limits, whereas the FF group is lower in Upper and Lower Limit. The HS group and the

Middle group have about the same distribution at upper and lower limits, whereas the FF group is lower in upper and lower limits than either of the other groups.

Turning now to the relationship with the TAT measure of n Achievement let us first examine the relationship between the HS-FF continuum and the total n Achievement score. Table 3 presents an analysis of variance of this threefold relationship. It is quite clear that the Middle group has the highest scores with HS and FF groups scoring at about the same level. In previous work persons with high scores had showed HS characteristics, with moderate or low scores being associated with signs of fear of failure. The present results are in obvious disagreement.

With respect to subcategories in the measurement of n Achievement, the McClelland scoring system includes several clearly positive and negative categories, which are given below. The positive categories represent what McClelland has earlier termed "Goal Imagery"; the negative categories represent what he has called "Deprivation Imagery."

Goal Imagery

Positive Categories

- I+ Successful Instrumental Activity
- GA+ Anticipations of Success
- G+ Positive Affect over Goal Achievement

Deprivation Imagery

Negative Categories

- I- Unsuccessful Instrumental Activity
- GA- Anticipation of Failure
- G- Negative Affect over Failure to Achieve Goal
- B Blocks or Obstacles in Path to Achievement

The following analyses involve the use of positive and negative subtotals computed from the scored protocols of each individual.

Table 4 presents an analysis of variance for both the positive and negative subscores. The Middle group employs significantly more negative categories than either the HS or FF groups. In regard to the positive subscores the three group means are very similar. The group differences in total n Achievement score are thus accounted for mainly in terms of the negative categories.

Because the Middle group has a much higher number of achievement-related stories, however, this group should tend to show a

TABLE 4

POSITIVE AND NEGATIVE N ACHIEVEMENT SUBSCORES FOR THE HS, MIDDLE, AND FF GROUPS (F_1 = between groups/within groups; F_2 = middle and combined HS, FF group/within groups)

Groups	Mean n Achievement Subscores		
	N	Negative	Positive
HS	14	1.78	1.43
Middle	16	3.81	1.37
FF	15	1.87	1.47
	$F_1 = 4.22$ $p < .025$	$F_2 = 8.43$ $p < .01$	$F_1, \text{ n.s.};$ $F_2, \text{ n.s.}$

TABLE 5

MEDIAN ANALYSIS OF HS-FF IN RELATION TO FREQUENCY OF POSITIVE AND NEGATIVE CATEGORIES

Group	Number of Subjects above or below Median with respect to:			
	Negative Categories		Positive Categories	
	Above median	Below median	Above median	Below median
HS	6	8	10	4
Middle	11	5	4	12
FF	11	9	9	6
	$\chi^2 = 3.13$ $p = .20$		$\chi^2 = 7.41$ $p = .03$	

* χ^2 was calculated using the method of maximum likelihood.

higher frequency for all categories. To take account of this difference the total positive and negative scores for each individual were divided by the number of his achievement-related stories. Table 5 presents the results of a median test for the number of positive and negative categories per achievement-related story. The median test is appropriate to the present data since it makes no assumptions about the variances, which would be dubious in the case of ratio scores. These results show that the HS and FF groups are about equal in relative frequency of positive categories, but are significantly higher than the Middle group. The findings for relative negative frequency are just the reverse except that the differences are not highly reliable.

DISCUSSION

Let us first examine the treatment of this general problem offered by McClelland (2).

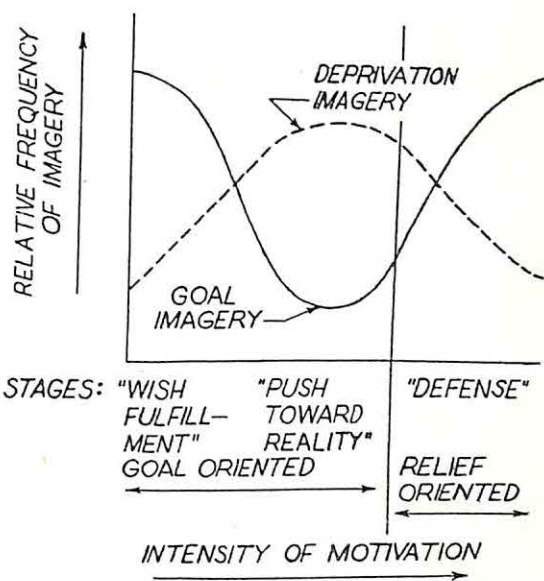


FIG. 1. THE HYPOTHETICAL EFFECT OF INCREASED MOTIVE INTENSITY ON THOUGHT PROCESSES

Reproduced by permission from McClelland (2). Copyright by William Sloane Assoc.

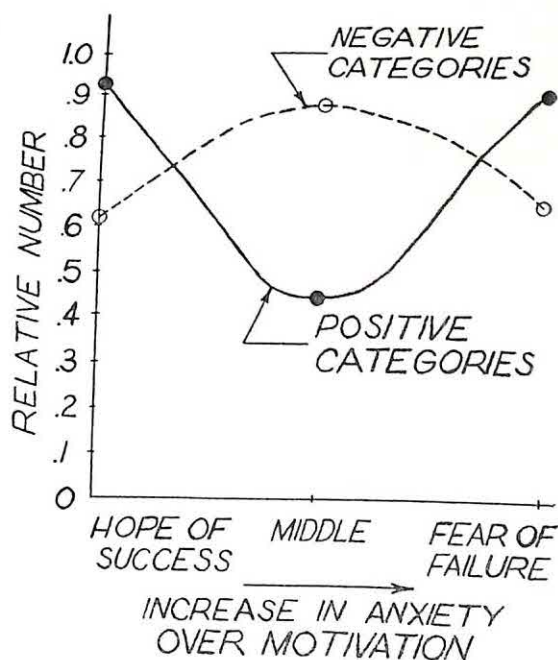


FIG. 2. THE NUMBER OF POSITIVE AND NEGATIVE CATEGORIES PER ACHIEVEMENT-RELATED STORY AS A FUNCTION OF THE HS-FF CONTINUUM.

Because there were five negative categories and only three positive categories used in this analysis, the number of positive categories has been multiplied by five-thirds in order to make the two curves directly comparable.

After a thorough review of the evidence related to the effects of different levels of intensity of motivation, McClelland presents the hypothetical functional relationships which are reproduced in Fig. 1. It can be seen that these curves are very similar to those in Fig. 2 which represent a plot of the results obtained from the analysis of the number of positive and negative categories per achievement-related story. McClelland's "Wish Fulfillment" stage, which corresponds to our HS group, represents a condition of satisfaction and no anxiety over failure to attain the goal in question. His "Push Toward Reality" stage corresponds to what we have called the Middle group. Here there is decreasing optimism and satisfaction and a moderate increase in anxiety over goal attainment. His final "Defense" stage is analogous to our FF group. Here anxiety is so great and hope of goal attainment so small that the individual must repress deprivation-imagery stimuli and enhance goal-related imagery. The interested reader should consult McClelland (2) for the complete treatment of this problem.

One general implication strongly suggested by this study is that the level of *n* Achievement score *cannot* be used as an index of the FF-HS continuum. It is quite probable that the total *n* Achievement scores obtained by groups along this continuum are a function of the pictures used to obtain the *n* Achievement scores. The discrepancy between the present results and earlier results could easily result from the fact that of the eight pictures used in this study, only one was the same as that used in the earlier work.

The present research also indicates why some of the earlier work failed to distinguish fear of failure and hope of success as aspects of achievement motivation. The present research shows that FF subjects are *phenotypically* similar to subjects characterized by Hope of Success. With regard to the TAT measure of *n* Achievement the total score and

the positive and negative subscores are similar for both of these extreme groups. It is the inclusion of the Middle group that gives rise to significant differences. In the initial attempt to arouse hope of success and fear of failure experimentally, successful arousal would lead only to the two extreme groups. Since no Middle group was provided for, the expected differences could not be found.

SUMMARY

Previous research has suggested that McClelland's TAT measure of *n* Achievement involves two recognizable aspects—hope of success (HS) and fear of failure (FF). A level-of-aspiration questionnaire related to grades on a course examination was designed to give an independent measure of the HS-FF continuum, and administered to college students in conjunction with the examination. Relations between this measure and both total *n* Achievement score and various *n* Achievement subscores were investigated. In brief, the findings were these: (a) students at the extremes of the continuum have lower *n* Achievement scores than students in the middle of the continuum. (b) in terms of positive subscores (goal imagery) on the *n* Achievement measure, the extremes have higher scores than subjects in the middle of the HS-FF continuum. (c) In terms of negative subscores (deprivation imagery) the Middle group tends to score more highly than the extremes.

REFERENCES

1. CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, 44, 391-399.
2. McCLELLAND, D. C. *Personality*. New York: Sloane, 1951.
3. McCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
4. MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.

Received September 22, 1955.

SEXUAL SYMBOLISM AND THE VARIABLES OF SEX AND PERSONALITY INTEGRATION¹

AUSTIN JONES²

Central State Hospital, Indianapolis

THE psychoanalytic theory of sexual symbolism has been notably lacking in experimental support since its formulation by Freud in his exposition of dream analysis (3). The theory assumes a symbolic relationship between the male and female sex organs and particular classes of objects on the basis of a functional or structural similarity between the organ and that which symbolizes it. The penis is thought to be symbolized by elongated, pointed, or penetrating objects, and the vagina by hollow, rounded, or enclosing objects. Although experimental verification of the theory is lacking, it has had rather widespread acceptance in certain areas of psychology and psychoanalysis and has become an integral part of the literature of projective techniques (1, 6).

A recent experimental study designed to test the hypotheses of sexual symbolism was reported by Levy (5). The specific hypotheses tested were (a) that Ss, when asked to pair symbols with male and female given names, would pair hypothesized "male" and "female" symbols with like-sex names more frequently over the total set of matchings than would be expected by chance, and (b) that the learning of "like-sex" pairs of names and symbols would be greater, as measured by retention scores, than the learning of "opposite-sex" pairs. The results of the experiment failed to support either hypothesis.

There are several factors in Levy's design, however, which confound the variables being studied and result in the experiment's lacking conclusiveness. The use of given names as part of the pairing task would seem to permit an unspecified range of earlier learnings to influence Ss' responses unless names are first equated for all relevant stimulus attributes. A

further source of limitation was the use of fifth grade children as Ss, most of whom are presumably prepubescent. It would appear that the degree of sexual maturity might be a factor relevant to the formation of sexual-symbolic responses.

The purpose of the present experiment was to reassess the hypothesis of sexual symbolism, utilizing as stimuli figures similar to those presented in Levy's study. The primary hypothesis being tested may be stated specifically as follows: When asked to designate stimulus objects as male or female, Ss respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female.

A second hypothesis, concerning the differential response of groups of varying degrees of personality integration, is proposed. Since the occurrence of symbolic responses is customarily regarded as a function of socially acquired inhibitions attached to more directly sexual responses, one would expect severe personality disorder, with its disorganization of social-inhibitory processes, to be accompanied by a less frequent occurrence of sexual-symbolic responses than is the case with comparatively well-integrated individuals. Thus, the second hypothesis of this study holds that hospitalized psychiatric Ss respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female less frequently than nonpsychiatric Ss.

METHOD

Subjects. There were two groups of 20 Ss each. The nonpsychiatric group consisted of ten male and ten female students in an introductory college psychology course. The psychiatric group consisted of ten male and ten female patients at a state mental hospital. Psychiatric and nonpsychiatric Ss were selected on the basis of age and education in an effort to equate the two groups as closely as possible on those variables. The mean ages for the nonpsychiatric and psychiatric groups were 22.0 years and 27.4 years respectively. The mean educational levels (highest grade completed) were 12.8 and 12.5. For the psychiatric group, patients grossly out of contact, with organic involvement or mental deficiency, and those having received insulin or electroshock treatment within the past six weeks were excluded as Ss. Schizophrenic disorders were predominant.

¹ The author wishes to express his appreciation to Dr. C. L. Williams, Superintendent of Central State Hospital, for permission to utilize patients as research subjects and to Dr. Morton Wiener, Chief Psychologist, for his advice and encouragement in the execution of this study. The author is indebted to James Norton of the Statistical Laboratory, Purdue University, for statistical help and advice.

² Now at The University of Rochester.

Materials. The stimuli in this experiment were the ten figures of hypothesized sexual symbolism used by Levy. Five are elongated, pointed, or penetrating ("male") and five are hollow, rounded, or enclosing ("female"). The figures are essentially "abstract" or without specific object identity; they are variations of circles, pyramids, rods, cubes, etc. In the present study, the figures were enlarged and reproduced in black ink on individual white cards.

Procedure. The Ss were seen individually by the author. The experimental procedure was third or fourth in a series of other experimental procedures, so there was ample time for the establishment of optimal rapport. The Ss were presented the following instructions verbally:

You know, we often think of things other than human beings as having personality. Very often these will be objects around the house or other things that are familiar to us. Some things seem to have a masculine kind of personality and some seem to have more a feminine personality. I have a set of cards here which have designs on them. I'm going to show them to you one at a time and you tell me which they seem to have more—a feminine or masculine personality. Just use your imagination and tell me your first impression. Don't try to think about it too much. Give the very first answer that comes to mind.

Following the instructions, *E* presented the ten figures in an order randomized for each *S*. The Ss were strongly urged to respond immediately. The few Ss who persisted in taking longer than two seconds per response were excluded from the study. This procedure was adopted in order to reduce the degree of conscious "problem-solving" types of responses and to prevent defensive, inhibitory responses. It has been held generally that the requirement of immediate response elicits responses that are less guarded or subject to repressive processes and that are in some way more "basic" to the individual's psychological functioning.

RESULTS

Table 1 summarizes the analysis of variance for the three variables of sex, psychiatric classification, and symbol type ("male" or "female"). Scores were the proportion of "correct" responses out of five for each symbol type—i.e., responses supporting the hypothesis of sexual symbolism. The arc-sine transformation for proportions was employed (7).

Comparison of the actual mean with the arc-sine equivalent of a chance expectancy of 50 per cent by *t* test indicated that Ss' responses were consistent with the hypothesis of sexual symbolism to a degree significantly different from chance expectancy ($t = 27.51, p < .0001$). Significant differences between sexes and between psychiatric classifications were found ($p < .05$ in each case). Differences between symbol types failed to reach significance, as was true of the various interaction effects.

TABLE 1
ANALYSIS OF VARIANCE OF PROPORTIONS OF "CORRECT" RESPONSES FOR EACH SYMBOL TYPE
(Data in the form of arc-sine transformation for proportions)

Source	df	MS	F
Between sexes	1	655.51	4.54*
Between psychiatric classifications	1	931.61	6.46*
Sex \times psychiatric classification	1	2.11	
Between Ss within sex \times psychiatric classification cells	36	144.23	
Between symbol types	1	132.61	
Sex \times symbol type	1	13.61	
Psychiatric classification \times symbol type	1	285.01	1.65
Sex \times psychiatric classification \times symbol type	1	74.11	
Symbol type \times Ss within sex \times psychiatric classification cells	36	173.64	

* Significant at the .05 level.

TABLE 2
ANALYSIS OF VARIANCE OF PROPORTIONS OF Ss GIVING "CORRECT" RESPONSE TO EACH SYMBOL
(Data in the form of arc-sine transformation for proportions)

Source	df	MS	F
Between sexes	1	7129	5.85*
Between psychiatric classifications	1	9486	8.08*
Sex \times psychiatric classification	1	360	
Between symbols	9	5158	4.24*
Sex \times symbol	9	1111	
Psychiatric classification \times symbol	9	1828	1.50
Sex \times psychiatric classification \times symbol	9	1218	

* Significant at the .05 level.

In order to assess between-symbols variation, an additional analysis of variance was performed and is summarized in Table 2. Data were in the form of the arc-sine transformation of the proportion out of ten Ss who gave the "correct" response to each symbol. The sex \times psychiatric classification \times symbol interaction was used as the error variance on the assumption that no interaction of this order really occurs. As before, the analysis shows significant differences for the main effects of sex and psychiatric classification. Significant variation between symbols is also shown ($p < .05$). Comparison of pairs of symbols was made by the method of Duncan and Bonner (2). In this analysis, one of the "female" symbols was shown to elicit "correct" responses to a signifi-

cantly lesser degree than all others. None of the other symbols showed significant differences from any other. Thus, the significant variation among symbols is accounted for by this particular one. It is interesting to note that the symbol in question, which might be described as a simple "box" or "cube" is the only one of the ten that is categorized as a sex-symbol on the basis of functional rather than structural similarity to the appropriate sex organ. All other symbols may be regarded essentially as instances of primary stimulus generalization.

Although differences between symbol types and also the symbol type \times psychiatric classification interaction failed to show significance in the analysis of variance, a comparison by *t* test of the mean scores for the psychiatric classifications for each symbol type revealed significantly higher scores for the nonpsychiatric group on "male" symbols than for the psychiatric group. The corresponding difference for "female" symbols was not found to be significant. The presence of the "anomalous" symbol noted above may be responsible for this outcome, however, for when recomputed omitting data from that symbol, $t = 1.92$ with $p < .07$.

DISCUSSION

The results of this experiment strongly support the first hypothesis, that Ss designate elongated, pointed, or penetrating objects as male, and hollow, rounded, or enclosing objects as female. The results are interpreted as support for the Freudian theory of sexual symbolism.

While the design of the experiment would seem to make plausible a moderate degree of generalization to the subject population, the degree of generalization to other stimulus situations is less clear. It would seem on safe and customary theoretical ground to assume that "response on the basis of sexual symbolism" in this experiment must have occupied some position in the Ss' response hierarchies prior to the experiment; otherwise, such responses could not have occurred. However, the general dominance of such responses in the Ss' response hierarchies is not established by this study. In other words, sexual symbolism may be regarded as having been demonstrated in the specific stimulus situation of this experiment and would be expected to generalize to all

stimulus situations embodying a significant degree of commonality with the experimental situation. The properties of the generalization gradient, however, remain unexplored. For this reason, application of the results of this study to more complex (e.g., clinical) situations is greatly restricted.

The second hypothesis, that psychiatric Ss respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female less frequently than nonpsychiatric Ss, was also supported. It appears that severe impairment of personality integration is accompanied by a disorganization of symbolic processes. As noted before, a significant difference was shown by *t* test between the psychiatric and nonpsychiatric groups in their response to male symbols. A similar difference was not found for female symbols. This finding was most likely due simply to the operation of the one "anomalous" female symbol. Thus, one can only report the result of significant differences between psychiatric and nonpsychiatric groups.

The finding of significant differences between sexes, with males responding more frequently on the basis of symbolism, seems to suggest that males are in some way more sensitized to sexual material and that they may be subject to less defensiveness than females in situations with sexual cues. Such hypotheses are consistent with the formulation advanced by Kinsey and his colleagues that "... males are more often conditioned by their sexual experience, and by greater variety of associated factors, than females" (4, p. 649). They report specific data indicating that males are more readily aroused sexually by a variety of visual stimuli than are females.

Some areas for subsequent research are suggested. The present study offers support only for sexual symbolism of a primary stimulus generalization type. Symbolism predicated on functional rather than structural similarity between the symbol and the organ symbolized remains unverified. A second area of needed research is the investigation of various subject variables, including period of sexual maturity and physiologic characteristics of sexual functioning, cultural determinants, etc.

SUMMARY

The present experiment attempted to investigate the hypothesis derived from the psy-

choanalytic theory of sexual symbolism that Ss, when asked to designate objects as male or female, respond to elongated, pointed, or penetrating objects as male, and to hollow, rounded, or enclosing objects as female. The experimental design involved the presentation of ten abstract figures of hypothesized sexual symbolism to psychiatric and nonpsychiatric Ss of both sexes. The figures, five "male" and five "female," were adapted from those employed in an earlier study by Levy, which failed to support the theory of sexual symbolism.

The total Ss responded consistently with the hypothesis to a degree significantly greater than chance expectancy, thus supporting the theory of sexual symbolism. It was noted that, while the theory was supported, the extent of generalization of the results of this experiment to other stimulus situations remains unspecified. Males were found to respond consistently with the hypothesis to a significantly greater extent than females, and nonpsychiatric Ss to a significantly greater extent than psychiatric Ss. None of the interaction effects proved sig-

nificant. Further research regarding symbolism predicated on functional rather than structural similarity of the symbol to the organ symbolized is particularly indicated, as the symbols of this study, with one exception, were of the structural type.

REFERENCES

1. BUCK, J. N. The HTP technique: a qualitative and quantitative scoring manual. Part one. *J. clin. Psychol.*, 1948, 14, 317-396.
2. DUNCAN, D. B., & BONNER, R. G. Simultaneous confidence intervals derived from multiple range and multiple F tests. Report presented to Amer. Statist. Ass., Montreal, Sept., 1954 (mimeographed).
3. FREUD, S. *The basic writings of Sigmund Freud*. New York: Modern Library, 1938.
4. KINSEY, A. C., POMEROY, W. B., MARTIN, C. E., & GEBHARD, P. H. *Sexual behavior in the human female*. Philadelphia: Saunders, 1953.
5. LEVY, L. H. Sexual symbolism: a validity study. *J. consult. Psychol.*, 1954, 18, 43-46.
6. MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.
7. WALKER, HELEN, & LEV, J. *Statistical inference*. New York: Holt, 1953.

Received October 13, 1955.

THE EFFECTS OF THREE VARIABLES ON CHILDREN'S CONCEPTS OF PHYSICAL CAUSALITY¹

MARTIN L. NASS

Educational Clinic, Brooklyn College

IN RECENT years, a number of studies, stimulated by some of the early work of Piaget (7, 9), have been devoted to a consideration of children's concepts of causal relationships. Piaget's "clinical method" of investigation, as he terms it, made use of individual interviews with children. A record was kept of the child's reply to questions regarding the mode of operation of physical phenomena. Seventeen distinct types of causal explanation, each characteristic of one of three different stages in the development of causal thinking, were distinguished. These stages range from finalism and magic, through animism, to logical deduction. Piaget regards the stage of a child's causal thinking as a reflection of his mental maturity and therefore as related to his chronological age. By the age of seven or eight years the child comes to distinguish between self and outer world and is more realistic in his approach to problems. It is at this age that the more rational types of causal explanation begin to appear.

Piaget's work was extended by other investigators to a consideration of factors other than chronological age on the development of causal thinking. Thus, Oakes (6), Granich (2), and Grigsby (3) considered the differences in causal thinking of children of varying IQ; Deutsche (1) dealt with the effect of socioeconomic status; Jones and Arrington (5) with race, and Williams (14) investigated the influence of familiarity with the problem situation in causal reasoning. For a critical review of the literature in this field up to 1943, the reader is referred to Huang (4).

The study reported in this paper analyzes the influence of three factors on the responses of children to various questions regarding

physical causality—personality, experience, and question form. The first and third of these factors have not been considered in previous research in this area; the second has been studied to some extent, but in a manner which differs from the present approach (14). The specific questions which the present study proposed to investigate are the following:

1. Does the personality of the child play a role in his concepts of causal relationships?
2. Does the nature of experience which the child has had with the phenomena about which he is questioned affect his ideas of how they operate?
3. Does the form or wording of questions regarding physical causality have an effect on the child's responses?

Personality. No study using Piaget's technique has dealt systematically or specifically with the influence of personality differences on causal concepts, although several investigators (6, 12) speculate on the significance of this factor. It is plausible to suppose that characteristic differences in basic modes of causal thinking exist between socially and emotionally normal children and children with definite tendencies to withdrawal and social avoidance. Often withdrawn children are likely to perceive the environment as threatening, with the anticipated result that "prelogical" modes of thought should predominate over naturalistic ones. In the present study questions were asked of both withdrawn and emotionally normal children to test the hypothesis that the causal thinking of withdrawn children is at a significantly less mature level than the causal thinking of normal children.

Experience. Systematic study of the child's degree of familiarity with the phenomenon he is questioned about has not been attempted. Studies that discuss this problem (2, 3, 6, 12) have raised certain questions regarding the definition of the term "familiarity." Ordinarily, the amount of experience (or frequency of encounter) the child has had with a phenomenon has been taken to repre-

¹ Based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Graduate School of Arts and Science, New York University. A more detailed report is presented in the original manuscript which is also available from University Microfilms, Publication No. 11, 946, \$1.35. The author wishes to express his appreciation to Professor Elsa E. Robinson, chairman of the committee, for her guidance throughout the course of this study.

sent his familiarity with it. However, results have shown that items concerning events of weather and movements of celestial bodies produce the more primitive types of causal thinking. The present study has adopted a different criterion for "familiarity" in order to circumvent this difficulty. "Familiarity" as considered herein refers to the child's possible experience with the actual agency of causality of the phenomenon.

A phenomenon can be classified into one of two experience levels, as follows: (a), Questions dealing with phenomena in respect to which the child himself can have had, and is likely to have had, direct experience with the agency or mechanics of causation. (b), Questions dealing with phenomena with which the child cannot have experienced the causal process directly. By shifting emphasis from the phenomenon itself to the causal processes underlying it, one can readily make this differentiation. It is then possible to test the hypothesis that questions concerning phenomena, the causal agents of which are less accessible to direct experience, yield more nonnaturalistic responses than questions concerning phenomena whose activators are more accessible to the child. In the present study, phenomena of weather and celestial bodies were chosen for the more "remote" experience level since the causal processes underlying them are outside the child's experience while the phenomena themselves are familiar and meaningful to the child. Furthermore, one would expect that the less logical explanations would be given to such items on the basis of the responses to similar questions in previous studies (1, 6).

Question form. Oakes (6) has suggested that the manner in which the question is worded can influence the type of response obtained. Piaget (8) also was aware of the influence of the verbal form of the question. He indicates that a wording form such as, "What makes the sun move?" may suggest the operation of naturalistic forces, and may mask any spontaneous tendencies toward animistic thinking. In Piaget's earlier studies on language (7) he points out that the child's early "Why" questions seem more affective than intellectual in character, and that they generally derive from a disappointment produced by the absence of a desired person or object, rather than signifying verbal curiosity.

A large proportion of the questions asked in previous studies including some of those of Piaget (8, pp. 291, 324), have begun with the word "Why." It is quite possible that such questions direct the child's thinking toward seeking some sort of purpose or motivation. It is one of the purposes of the present study to test the hypothesis that questions worded so as to suggest the possible operation of animistic or dynamic forces yield more non-naturalistic responses than those which are not so worded.

METHOD

The subjects of the study were 120 children, aged 8-0 to 10-0, enrolled in Grades III and IV of the New York City public schools. Sixty of the children were "emotionally disturbed" and sixty of them were of "normal" emotional adjustment. The disturbed groups consisted of children whose initial referral problems involved "withdrawal." They were selected on the basis of records of the Bureau of Child Guidance of the New York City Board of Education. Normal children were selected on the basis of their over-all school adjustment as noted on the school record and health cards and as judged by their classroom teachers. No child who had been known to a social agency because of adjustment problems was included in the normal group.

The normal and disturbed groups were selected and matched with regard to age, sex, and intelligence test scores. Wherever possible, the technique of matching entailed the selection for the normal group of a classmate of a withdrawn child. All children were of average intelligence (IQ 90-110) as determined by the Pintner-Cunningham Intelligence Test administered in the first grade.

Each of the major groups was divided into two equal subgroups, similar in composition with regard to age, sex, and intelligence-test scores. The experimental design and the composition of the groups with respect to the matching variables are given in Table 1.

Each child was seen individually by the investigator and, after being put at ease in the situation, was given the following instructions:

We are going to play a game. I am going to ask

TABLE 1
COMPARISONS OF THE GROUPS WITH REGARD TO
MATCHING VARIABLES

Variable	Group I	Group II	Group III	Group IV
Wording form	A	B	A	B
Adjustment	Normal	Normal	With- drawn	With- drawn
Experience level	1, 2	1, 2	1, 2	1, 2
Mean age	9.2	9.0	9.0	9.0
SD (age)	0.9	0.7	0.7	0.7
Mean IQ	100.5	99.9	100.6	98.4
SD (IQ)	9.2	8.5	9.5	9.5

Note.—Each group is composed of 15 boys and 15 girls.

TABLE 2
LIST OF QUESTIONS USED IN THE STUDY

Experience Level	A	B	
I	(Why does)	(How does)	the radiator get hot?
	(Why do)	(How is it that)	boats stay on top of the water?
	(Why does)	(How is it that)	a clock ticks?
	(Why does)	(How does)	a whistle blow?
	(Why does)	(How does)	a car move?
	(Why do)	(How come)	leaves fall off the trees?
	(Why do)	(How is it that)	balloons go up in the air?
	(Why do we get shadows?)	(How do shadows come about?)	
II	(Why do we have waves?)	(How do waves come about?)	
	(Why do)	(How do)	we get thunder?
	(Why do)	(How is it that)	we have wind?
	(Why do)	(How is it that)	the stars shine?
	(Why do)	(How is it that)	clouds move?
	(Why do)	(How is it that)	we sometimes have rainbows after the rain?
	(Why does)	(How come)	the sun shine(s) on some days but not on others?
	(Why does night come?)	(How is it that we get night?)	
	(Why is the sky blue?)	(How does the sky get blue?)	
	(Why do we have snow?)	(How does snow come about?)	

you some questions and we will see how many you can answer. Some of these questions are pretty hard, so don't worry if you can't get many. This has nothing to do with your school work and your teacher won't know how you do. I'm just trying to find out what boys and girls know about different things.

The child was then asked the questions in either Form A or Form B of the questionnaire. The materials covered in both forms were selected largely from previous studies and were identical except for the fact that the questions of Form A were posed as "Why's" and therefore might be more suggestive of the operation of dynamic or animistic forces than the questions of Form B, each of which started with "How." The questions are listed in Table 2. Questions from the two experience levels were presented in mixed order. Responses were recorded verbatim, and ambiguous responses were probed by such neutral questions as, "What do you mean?" or "How does that happen?"

Each response was scored independently by three judges.² The materials were arranged so as to prevent identification by the judge of either the wording form of the question or the adjustment class of the subject. Responses were classified according to the following scheme:

1. *Nonnaturalistic*. This category included all of Piaget's "precausal" stages which are nonmaterialistic, i.e., responses containing causal statements relating to spirits, magic, or other nonmaterialistic phenomena. It included his motivational, finalistic, moral, magical, animistic, and dynamic classes. Previous studies have shown that areas of overlap among Piaget's 17 cate-

gories are so great that it is often virtually impossible to classify a given response definitively (Cf. 4, p. 73).

2. *Phenomenistic*. This is also a "precausal" type, but may or may not be of a materialistic nature. Phenomenistic causality is described by Piaget as follows:

Two facts given together in perception, and such that no relation subsists between them except that of contiguity in time and space, are regarded as being connected by a relation of causality (9, p. 259).

3. *Naturalistic*. This category included responses which are essentially naturalistic and materialistic. This included Piaget's mechanical and logical categories. Responses need not necessarily be correct to be placed in this category, but must contain underlying ideas which are naturalistic.

RESULTS

Agreement among the judges on categorization of responses was unanimous in 95.7 per cent of the responses. In 3.9 per cent of the responses, agreement in two out of three judgments was obtained. The remaining responses were excluded from analysis. Listed below are examples of responses and their scoring.

Question 1. "Why do leaves fall off the trees?"

a. "Because when winter's coming they couldn't stay on. If they stayed on the snow wouldn't look nice. If they stayed on, winter

² The judges were Professor Elsa E. Robinson, Mr. Libero Arcieri, and the author.

TABLE 3
FREQUENCY OF NONNATURALISTIC RESPONSES FOR
COMBINED GROUPS

Basis of Grouping	Percentage of Nonnaturalistic Responses	<i>T</i>	<i>p</i>
Normal vs. with- drawn children	18.9 vs. 45.6	2543	<.001
Familiar vs. remote causal agency	19.3 vs. 45.7	126	<.001
"How" wording vs. "why" wording	15.3 vs. 49.3	2293	<.001

Note.—While the *T* tests were based on rankings of the number of responses in the nonnaturalistic category, percentages are presented here for more meaningful comparison.

TABLE 4
FREQUENCY OF NONNATURALISTIC RESPONSES
FOR SUBGROUPS

	Percentage of Nonnaturalistic Responses	
	Normal	Withdrawn
Causal agency "familiar"		
"Why" wording	14.8	54.9
"How" wording	0.4	7.4
Causal agency "remote"		
"Why" wording	48.8	78.5
"How" wording	12.2	43.3

Note.—By Wilcoxon *T* tests, all four relevant comparisons for each basis of grouping (normal vs. withdrawn, familiar vs. remote, "why" wording vs. "how" wording) are significant at the .001 level.

couldn't come." Scored "nonnaturalistic" by all three judges.

b. "Cause it's cold. They just fall off because it's cold. Because it's winter." Scored "phenomenistic" by all three judges.

c. "They turn different colors and dry up and fall off because they die." Scored "naturalistic" by all three judges.

Question 8. "How is it that we sometimes have rainbows after the rain?"

a. "It's a sign to tell you that the rain is over." Scored "nonnaturalistic" by all three judges.

b. "The sky makes it. Rainbows are in the sky, so the sky makes it." Scored "phenomenistic" by all three judges.

c. "The sun is red and the sky is blue and the clouds are white, so they all come together and you see all the colors together

and that's a rainbow." Scored "naturalistic" by all three judges.

Since the "phenomenistic" category accounted for only six per cent of the total responses, analysis was confined to the number of responses falling into the "nonnaturalistic" category, to which the "naturalistic" category was thus essentially the complement. The comparisons were made using Wilcoxon's *T* tests (13) since the distributions of responses differed significantly from normality and from each other. The results of the comparisons are presented in Tables 3 and 4.

The effects of each factor are clearly shown when the responses of a given group under different conditions are compared. Thus, normal children responding to materials of a more familiar nature which are presented in a nonsuggestive manner give on the average 0.4 per cent nonnaturalistic responses. With less familiar materials this value rises to 12.2 per cent. However, when suggestive wording is introduced the proportions are 14.8 per cent and 48.8 per cent, respectively. Similar differences at a higher level of frequency are noted in the withdrawn groups.

Thus, the causal responses of children to verbal questions concerning physical causality were significantly influenced, as hypothesized, by (a) the personality differences tested in this investigation, (b) the contact which they may have had with the agency of causation of the phenomena, and (c) the wording of the question.

DISCUSSION

Although Piaget does not specifically mention the possible influence of personality factors on level of causal thinking, some of his writings discussing the development of thought processes in children help to clarify his views regarding this variable. Piaget relates the development of more mature thinking to the child's ability to differentiate his subjective world from the external world. During the early stages of development the world and the self are one. Gradually the realms become more distinct, conceptually, until in the final stage the child comes to distinguish between the subjective and objective points of view. This decrease in egocentrism is related to decrease in animism, in anthropomorphic finalism, and in feelings of participation in

external phenomena. The child begins to conceive of reality as constituted by what is common to all points of view taken together (9). The results of the present study with regard to personality differences may be related to this position. The withdrawn child, who feels himself threatened by the environment, may be conceived in Piaget's terms as being hampered in developing the objective point of view, i.e., as more likely to respond syncretically where the emotionally healthier child clearly discriminates self from nonself.

The present findings concerning personality differences leave open the question as to whether the causal factor lies in specific withdrawal tendencies or general maladjustment. Since the "problem" group employed in the present study was restricted to children who emphasized one characteristic mode of handling the environment, i.e., withdrawal, it is not possible to state how a group of aggressive children would respond, or what the results of a comparison between a combined group of "disturbed" vs. "normal" children would be. There are grounds for expecting the withdrawn group, whose retreat is accompanied by fantasy and autism, to produce more nonnaturalistic types of responses than other maladjusted children, but it is not possible to test this hypothesis through the present design.

Experience with a phenomenon as a potential influence on the child's causal concepts is not directly mentioned in Piaget's writings. However, in his later writings on intelligence (10, 11) Piaget extensively discusses the importance of early experiences on later responses. It is not likely that Piaget would deny the importance of experience in determining the level of causal thinking while ascribing a major role to it in the genesis of intelligence. The fact remains, however, that in the volume he devotes to the subject of causal thinking in children (9), Piaget considers neither experimentally nor interpretatively the role of experience in the results which he reports. While the present results uphold the hypothesis that the level of causal thinking may be influenced by experience with the underlying causal agent of a phenomenon, it should be noted that the questions included in the level of more remote experience dealt only with meteorological phenomena.

There is a possibility that the differentiating factor may have been something peculiar to these phenomena themselves, rather than experience with their causal processes.

The possible influence of suggestive wording on the responses of children to questions which deal with causality was briefly considered by Piaget, although his exposition of the subject seems to have been logically rather than empirically derived. The present results can be interpreted as lending support to his speculations in this area. The fact that a subsidiary test³ showed the effects of suggestive wording to be more pronounced with remote than with familiar materials for normal children tends to support some of Piaget's later formulations. He says, "It is true that the more active experience is, the more the reality on which it bears becomes independent of the self and consequently 'objective' (11, p. 367)." Certainly, the more familiar items have been experienced more actively and hence the findings with regard to experience can be reconciled with Piaget's thinking on the subject.

Although these accommodations to Piaget's theory can be made, it should be pointed out that Piaget emphasizes almost exclusively the relation of level of causal thinking to chronological age. On the basis of his exposition, one would expect that most children between the ages of eight and ten years would offer mainly naturalistic causal explanations. The present results, however, show that for some questions the explanations range through most of Piaget's 17 stages at an age at which he contended most should occur near the culminating stages, and fail to support the view, strongly suggested in certain of Piaget's writings, that each developmental level is characterized by a peculiar conception of causality which is generally applied to all the phenomena which the child experiences.

Nevertheless, the present findings at the same time uphold Piaget's insistence that much of the causal thinking of children is prelogical. Although he contended that the naturalistic and logical forms of thinking began to appear by seven or eight years in

³ A comparison of the differences in number of nonnaturalistic responses between remote and familiar materials showed a significantly greater number for suggestively worded questions than for nonsuggestively worded questions in normal children ($T = 702.5$; $p < .001$).

his sample, and became prevalent in the explanations of the 10- and 11-year-olds, the present study has shown that for the objects and phenomena involved, many of the eight- to ten-year-old children entertained causal ideas which are nonlogical and nonnaturalistic some of the time, and some of the children entertained such ideas most of the time. The data indicate that the nature of children's causal thinking is affected by traits of personality, by the manner in which the question is worded, and by the subject matter of the question.

SUMMARY

This study investigated the effects of three variables on children's concepts of physical causality. It was found that:

1. The nature of the causal thinking of withdrawn children is at a significantly less mature level than the causal thinking of normal children.

2. Questions about phenomena whose causal agents are not accessible to direct experience yielded significantly more nonnaturalistic responses than did questions about phenomena whose causal agents are more accessible.

3. Questions worded so as to suggest the possible operation of "animistic," "supernatural," or "dynamic" forces yielded more such nonnaturalistic types of responses than questions less suggestively worded.

REFERENCES

1. DEUTSCHE, JEAN M. *The development of children's concepts of causal relations*. Minneapolis: Univer. of Minnesota Press, 1937.
2. GRANICH, L. A qualitative analysis of concepts in mentally deficient schoolboys. *Arch. Psychol.*, N. Y., 1940, No. 251.
3. GRIGSBY, OLIVE J. An experimental study of the development of concepts of relationships in pre-school children as evidenced by their expressive ability. *J. exp. Educ.*, 1932, 1, 144-162.
4. HUANG, I. Children's conception of physical causality: a critical summary. *J. genet. Psychol.*, 1943, 63, 71-121.
5. JONES, F. N., & ARRINGTON, M. G. The explanations of physical phenomena given by white and Negro children. *Comp. Psychol. Monogr.*, 1945, 18, No. 5.
6. OAKES, M. E. Children's explanations of natural phenomena. *Teach. Coll. Contr. Educ.*, 1947, No. 926.
7. PIAGET, J. *The language and thought of the child*. New York: Harcourt, Brace, 1926.
8. PIAGET, J. *The child's conception of the world*. New York: Harcourt, Brace, 1929.
9. PIAGET, J. *The child's conception of physical causality*. New York: Harcourt, Brace, 1930.
10. PIAGET, J. *The psychology of intelligence*. New York: Harcourt, Brace, 1950.
11. PIAGET, J. *The origins of intelligence in children*. New York: International Universities Press, 1952.
12. SARVIS, B. C. *A clinical study of the developmental aspects of child thought*. Unpublished doctor's dissertation, Univer. of Kansas, 1939.
13. WILCOXON, F. *Some rapid approximate statistical procedures*. New York: American Cyanamid Co., 1949.
14. WILLIAMS, M. *The influence of specific experiences on children's concepts of causal relations*. Unpublished doctor's dissertation, Univer. of Michigan, 1940.

Received November 3, 1955.

SPEED OF FORM RECOGNITION AS A FUNCTION OF STIMULUS FACTORS AND TEST ANXIETY¹

FREDERICK C. ROCKETT²

The Dow Chemical Company

DURING the past decade, studies have appeared in the psychological literature reflecting an increasing interest in the etiology as well as the manifestations of color shock, generally considered a sign of neuroticism, in the Rorschach test (18). The results of many studies suggest that certain inkblots seem to provoke "color-shock" responses whether the blots are colored or achromatic (1, 2, 6, 11, 12, 16, 17, 19). All of these studies, however, suffer from the lack of a proper control group representing neurotic or highly anxious Ss. Studies employing a proper control group have been more successful in supporting the hypothesized relationship between color, neuroticism, and certain responses to colored inkblots (3, 4, 13), but none has found more than tenuous evidence.

However, from the speculations of various students of color shock have come several conceptions regarding its etiology. Schachtel (21) postulates an innate responsiveness of neurotics to color, while Wallen (24) suggests that positive and negative affective responses to color are due to the association of color with reward and punishment respectively. Siipola (22) believes that hue-form incongruity or the complexity of the stimulus may greatly alter response content. A number of studies suggest that stimulus complexity is the main determinant of color-shock indices (5, 11, 17). In general, there are two approaches to the interpretation of color shock. One is through characteristics of subjects, and the other through stimulus characteristics.

The present study is an attempt to test various color-shock hypotheses and their underlying theoretical assumptions through an analysis of response time. The color-affect theory (Schachtel, Wallen) concerning color shock in the Rorschach test suggests that

highly anxious individuals have a relatively delayed reaction to brightly colored cards because of a negative affective response to color. Nonanxious individuals are comparatively unaffected by color and consequently have faster reaction times. The following hypotheses were framed for investigation:

Hypothesis 1. Highly anxious individuals take significantly longer to recognize colored ambiguous figures than less anxious persons.

Hypothesis 2. Recognition times for figures having disliked colors are significantly longer among highly anxious persons than for figures having pleasant colors.

Hypothesis 3. Disliked colors are more often linked with unpleasant concepts (e.g., blood) in highly anxious individuals than in less anxious ones.

Students of color shock who emphasize the role of stimulus complexity suggest that the relationships proposed in Hypotheses 1, 2, and 3 are less important than such factors as figure-ground contrast through the juxtaposition of various colors, the intrinsic difficulty of the figure, and other stimulus-complexity determinants.

Hypothesis 4. Stimulus attributes contributing to complexity contribute significantly more variance to total performance than anxiety level or color preference.

Cox and Sarason (4), like others (7, 23), postulate that anxiety has drive characteristics, while Miller (14) suggests that any strong stimulus may have drive characteristics. From Hullian behavior theory (10), it follows that with increasing drive strength, response time to simple stimuli becomes faster and suprathreshold responses more frequent and stronger in a free response situation. Thus, the degree of response competition, the number and strength of simultaneously elicited responses, is positively related to drive level and stimulus complexity. When response competition is strong, however, then response time should be faster for less anxious than for more anxious Ss in a situation in which only one response can be made. Godbeer (8) has shown that in such instances multiple approach-avoidance conflicts occur. It follows that re-

¹ A condensation of a dissertation presented in partial fulfillment of the requirements for the Ph.D. degree, University of Texas, 1955.

² The author is grateful to the faculty of Wichita University for their hospitality and aid in the completion of this study. Dr. N. H. Pronko deserves particular recognition.

sponse time should be positively related to the number and strength of these conflicts.

Hypothesis 5. Factors influencing drive level (anxiety, color preference) are negatively related to response time when stimulus complexity is low, but positively related to response time when complexity is high and only one response is allowed.

METHOD

Experimental design and materials. The hypotheses were tested in a single design in which a test series followed a warm-up series of four forms from the Holtzman Form-Recognition Test (9).

The Holtzman Forms are line drawings of common animals, birds, reptiles, and humans in which identification is delayed by omission of parts of the figure and by the addition of superfluous lines. The average response time to the figures selected for this study ranges from 1.10 seconds to 3.00 seconds. Correct identifications took place 95 per cent of the time in the original standardization with college students. The Holtzman Forms were used in order to reduce the difficulty encountered in many Rorschach studies of dealing with stimuli so complex that many responses are commonly given to a card. In such instances, it is often difficult to determine what is or is not a deviant response, and statistical problems of analysis are greatly complicated.

Warm-up cards were presented so that novelty would not delay responses, thereby contributing to the total variance in response time of the main study. A pilot study and the work of Doidge (5) showed that response times reach an asymptote with the administration of four practice cards. The order of presentation was the same for each subject, progressing from lesser to greater levels of difficulty with the most difficult warm-up figure being somewhat less difficult than the easiest test figure.

Colors used in the test series were selected for closeness to the four psychological primaries and for brightness. The Munsell designations of the colors are presented in Table 1. The patterns in which the colors were presented were selected to represent a wide range of possible patterns, to follow the work of Doidge, and to lend themselves to fairly uniform duplication. The four patterns were wave, speckled, splotched, and solid. Colors and patterns were combined by spray-painting each of the patterns in each of the colors on sixteen glass plates. Such plates, when placed before the Holtzman Figures, gave the appearance of colored patterns and figures on single surfaces. The wave and splotched patterns were stencilled for greater uniformity.

The viewer used was specially constructed, using the

TABLE 1
MUNSELL COLOR VALUES OF TEST SERIES

Hue	Lightness	Chroma
Greenish-yellow (7.5 Y)	8	10
Red (5.0 R)	4	14
Blue (5.0 B)	4	8
Yellowish-green (2.5 G)	5	8

main component of the Harvard Tachistoscope as a model. It was $9\frac{1}{4}$ " square and $21\frac{3}{4}$ " long. The inside of the viewer was painted flat black on all surfaces except the areas around two recessed light sources, which were glossy white. The light from two long unfrosted bulbs of 25 watts was made diffuse by the interposition of frosted glass.

Of the several scales which purport to measure neuroticism and anxiety, only two have been utilized to such an extent that some confidence can be expressed in their utility. These two scales are the Taylor Manifest Anxiety Scale (23) and the Sarason Anxiety Scale (20). The Sarason was used rather than the Taylor because it appears relatively innocuous. The Ss were not given the idea that they were being evaluated critically, and were therefore not forced into a state of anxiety due to an adventitious situational factor.

An arabic-greco-latin square (AGLS) design was employed, composed of three independent latin squares superimposed on each other so that each of three independent factors appears once, and only once, in each row, each column, and in combination with each of the other two factors. Sixteen different AGLSs were necessary to represent all possible combinations of the three experimental stimulus-variables (figure, color, pattern) in the first column of the AGLS (Trial I). Since four Ss are needed to complete each square, 64 Ss constitute one complete design. The addition of three levels of anxiety as an independent variable required a total of 192 Ss. This design allowed for the testing of the interaction of the main variables with anxiety and AGLS, as well as allowing the testing of these variables individually.

The first response to the test series for each S was analyzed in a $4 \times 4 \times 4 \times 3$ analysis-of-variance design without replication. It was possible to carry out this analysis while maintaining orthogonality because each variable was presented once in the first trial of each square.

Subjects. The Ss were 192 male students attending Wichita University. The entire Air Reserve Officer Training Corps, all freshmen of the Army Reserve Officer Training Corps, and several classes in introductory psychology served as the population from which the sample was drawn.

Procedure. The Sarason Anxiety Scale (19) was administered during regularly scheduled classroom hours. All Ss falling in the middle 15 per cent and the two extreme 15 per cent segments of the distribution obtained on the Sarason Scale were eligible for the main study. Samples from the middle as well as the extremes were drawn because a previous study (7) had indicated a marked curvilinear relationship between performance and anxiety level. These men were called by telephone for individual appointments. Only two Ss refused to participate.

Having arrived at the experimental room, the S was seated before a viewer. The room itself was a 9-ft. \times 10-ft. windowless chamber containing only material relevant to the present study. The S was then given a color-blindness test using the Dvorine Color Charts. If S passed this test, E read the following instructions; otherwise the session was terminated at that point. Eleven Ss were discarded as color blind.

Look into the viewer. Presently a series of cards

will be shown to you. On each card there will be a figure, such as an animal, bird, insect, or human. It is your task to identify these figures as quickly as possible, using the specific name in each case, such as chicken, lizard, etc. I will be timing you with a stop watch.

Remain in position for the next card after the light goes off.

Are you ready?

While these instructions were being read, *E* prepared the stimulus materials for the first presentation. The first card of the warm-up series was inserted into an especially constructed holder attached to the end of the viewer opposite to the viewing aperture. A clear glass plate was put just in front of the figure to insure uniformity of illumination of the figures for both the warm-up and test series.

Immediately after the instructions were read, the first warm-up figure was presented by turning on the lights in the darkened viewer. When a response was made, the lights were turned off, the response time and response were recorded on a form, and the second stimulus prepared. A shield was placed in the position ordinarily occupied by the stimulus holder during changes of stimuli to prevent *S* from seeing the characteristics of the forthcoming material. The other stimuli followed in the same fashion without further communication with *S*, except to indicate that another figure was about to be presented. Care was taken to give *S* no clue to the shift to colored stimuli after the fourth warm-up card.

A 60-second time limit was allowed for making each response, after which time the light in the viewer was turned off and preparation for the next trial begun. If *S* wanted to give up before the 60 seconds had passed, *E* encouraged him to continue trying.

After the response times and responses were recorded, *S* was asked to stand before a table placed near the viewer. On the table were the four glass plates bearing the color-pattern combinations that *S* had seen during the test series. Under the plates were pieces of white paper of approximately the same reflectance as the photographic paper on which the figures were printed. The color-patterns then had an appearance similar to those which were seen in the viewer except that the figures were absent. At this time the order of preference and first association to the color patterns were recorded. Lastly, questions designed to assess the reliability and validity of the Sarason scale as a measure of neurotic anxiety were asked. They were:

1. The questionnaire that you took for me tried to get at how confident you are in taking tests. Your score on the questionnaire fell in the (15 per cent most confident, 15 per cent least confident, or middle 15 per cent). Do you think that this is reasonable?
2. How confident are you socially?
3. Have you ever had any kind of personal counseling?
4. What kind, and where?
5. How well adjusted do you feel? How well satisfied with yourself as a person are you?
6. To what extent do you think that you would fear a personality evaluation such as taking a battery of tests and being interviewed by a psychiatrist?

RESULTS

Responses to the above questions were categorized in three divisions for statistical purposes. These three divisions represent the extremes and the average of the range of responses. Chi square revealed that responses to questions 1, 2, 5, and 6 are highly related to the Sarason-scale scores ($p < .001$). That is, high anxiety as indicated by the Sarason scale was highly related to responses thought to indicate neuroticism. No difference in frequency of reported attendance to a counseling service was found for the three anxiety groups.

To normalize the typically skewed distribution of response-time scores, an inverse square-root transformation was applied. All statistical calculations involving response-time data utilized transformed scores.

The results of an analysis of variance of response time during warm-up trials showed significant differences between anxiety levels ($p = .01$), trials ($p = .001$), and the interaction between anxiety and trials ($p = .01$). The highly anxious group proved to be somewhat slower in response time than the other two groups, although they registered the fastest mean response time on the first trial. In each trial thereafter in the warm-up series, the high anxiety group became progressively slower relative to the medium and low anxiety groups.

Since figures were presented in the same order for all *Ss*, trials and figures are confounded. Nevertheless, because the warm-up figures were presented in an ascending order of difficulty according to standardization information, and because the median response time for each trial or figure became progressively shorter except for the last of the series, it is clear that there were strong practice effects which obscured the response-time characteristics of the figures.

The comparison of the mean difference score for the last warm-up and the first of the colored cards for the high and low anxiety groups, the two with the greatest mean difference, showed no significant difference. All groups displayed similar increases in response time, significant beyond the .001 level of confidence, to the first colored card.

Analysis of variance of response times to the first colored card revealed the powerful role played by the color ($p < .001$), figure

TABLE 2
ANALYSIS OF VARIANCE FOR THE ARABIC-GRECO-
LATIN SQUARE DESIGN

Source of Variation	df	Mean Square	F	F†
Independent Observations:				
Different AGLS	15	63.0	1.60	
Anxiety levels	2	75.0	1.90	
Anxiety × AGLS	30	36.0	.93	
Residual between Ss	144	39.0		
Total between Ss	191			
Correlated Observations:				
Trials	3	47.0	3.00*	1.6
Color	3	380.0	24.50***	
Pattern	3	966.0	62.30***	
Figure	3	632.0	40.80***	25.5***
Trials × AGLS	45	29.0	1.90**	
Color × AGLS	45	20.0	1.30	
Pattern × AGLS	45	19.0	1.20	
Figure × AGLS	45	25.0	1.60*	
Trial × Anxiety	6	20.0	1.30	
Color × Anxiety	6	12.0	.77	
Pattern × Anxiety	6	7.0	.45	
Figure × Anxiety	6	12.0	.77	
Residual	360	15.5		
Total within S	576			
Total for experiment	767			

* $p < .05$, ** $p < .01$, *** $p < .001$.

† These additional F ratios have been computed using the trial AGLS interaction and the figure AGLS interaction to test the significance of trials and figures, respectively.

($p < .001$), and color-pattern ($p < .001$) variables. Anxiety level did not influence response time to an appreciable degree. The interactions of pattern and figure ($p = .01$) and of color, pattern, and figure ($p = .05$) also played a significant role in response-time determination.

Table 2 summarizes the results of the analysis of variance for the AGLS design. Colors, patterns, and figures were found to be potent factors in the determination of response time while trials and the interactions of both trials × AGLS and figures × AGLS were significant but lesser contributors. Anxiety level did not play a significant role in response-time determination in the test series, either alone or in combination with other variables.

Analysis of color, pattern, and color-pattern preferences showed no difference among anxiety groups. Colors have a wider range of

preference frequencies than do the patterns, indicating a greater range of affectivity for color. Chi-square analysis of the frequency of each degree of preference for each color shows that colors tend to have significantly different ($p = .01$) frequencies at each degree of preference. Differences in frequency of preference for patterns were significant ($p = .01$) only for the most preferred category. Ranking colors in order of response time and comparing the order to that found for color preference indicated that the relationship between response time and preference is the exact reverse of the prediction from color-affect theory that negative affect to color is associated with longer response times.

Further analysis of response time for the three anxiety groups and for degree of color preference uncovered no significant relationship between anxiety group, color preference, and response time. Combining the anxiety groups did not alter the finding of no significant difference in response time between degrees of color preference.

Only a few associations to the color patterns were regarded as unpleasant. Some typical associations were linoleum, flowers, waves, and germs. There were only 5 "blood" associations in all 768 opportunities, which included 48 presentations of a red splotched pattern. This number was too small for statistical consideration in relation to anxiety groups.

In order to sharpen the anxiety variable, an especially selected maladjusted group was compared in several ways with an especially selected well-adjusted group. The Ss who had indicated maladjusted responses on any two of the three interview items of social anxiety, fear of psychiatric evaluation, or degree of adjustment, or who had indicated that they had had psychiatric treatment, were grouped and their various scores compared with those of the medium and low anxiety categories. This gave a total of 45 Ss in the maladjusted group to be compared with 109 Ss who neither met the above criterion nor scored in the highly anxious range on the Sarason Anxiety Scale. No evidence was found to indicate that this further refinement of the anxiety variable yielded results different from those reported.

DISCUSSION

Hypothesis 1, that highly anxious individuals have longer response times than those

less anxious, was not substantiated. The three anxiety levels did not have significantly different mean response times in the test series. Similarly, no differences were found among anxiety levels in response time between the last warm-up card and the first card of the test series.

Hypothesis 2, that disliked colors result in longer response times than liked colors for high anxiety Ss, was also unsupported. No significant difference in response time between color preferences for highly anxious Ss was found. It should be recognized, however, that several variables of known significance were uncontrolled in this test. Complete control of color, color pattern, and figure was not possible, and it is likely that some imbalance was present, particularly for color.

Strengthening the notion that color preference is not a major factor are the negative results of a pilot study in which color pattern, color preference, and figure were independent variables. These findings are consistent with evidence (14) indicating that color preferences are largely learned and are quite ephemeral, being amenable to rapid change. Further, Warner (24) exhaustively studied the color preferences of various psychiatric groups, finding little of consequence other than the usual male-female differences. It is likely therefore, that most persons are not affected by color preferences to the same degree that they are by other factors.

As an incidental observation and contrary to usual clinical belief the color red was neither disliked by more highly anxious persons than by others, nor was it associated with longer response time.

A weakness in the study is indicated by the lack of affect-laden associations to the color patterns. This defect vitiates considerably the testing of Hypothesis 3. However, it seems interesting, considering common practice, that the association "blood" was so rare an occurrence, despite the frequent use of a red splotched pattern.

Complication theory seems best supported by these results. As Doidge and others have indicated, factors contributing to stimulus complexity delay response time, whereas other factors remain relatively unimportant in this respect. Hypothesis 4 was confirmed by the powerful stimulus effects of colors, color pattern, and figure in response-time deter-

mination; while the influences of such respondent variables as color preference and anxiety were negligible. The significant difference between trials and the significant interaction between trials and AGLS cannot be dealt with by complication theory in its present form, but these findings represent order effects, not centrally relevant problems in the relationship of affect to color.

Hypothesis 5 was only partly confirmed because factors which theoretically could have altered drive level did not function consistently. Neither anxiety level nor color preference had any discernible effect on response time in the test series, although anxiety level and the interaction of anxiety and trials were significant determinants of response time in the warm-up series.

Inspection of the means for each trial at each anxiety level for the warm-up series suggested that response time for the highly anxious Ss became progressively slower relative to the medium and low anxiety Ss. Conceivably, a multiple approach-avoidance conflict may have been greater for high anxiety Ss. If this explanation were true, however, one would expect that the greater difficulty of the test series would increase the conflict, further slowing response time. This was not found. Thus, warm-up findings may have been due to chance alone, or it may be that some other explanation is possible. Future research could profitably explore greater ranges of drive level and stimulus complexity and their relationship.

It can be said that little confirmation for any explanation of color shock was found except for complication theory. The idea that response time is largely a matter of stimulus attributes and not the anxiety of Ss received strong support. However, because they generate testable hypotheses combining stimulus and subject characteristics, more general theories of behavior seem to offer the most promising route to a fuller understanding of responses to ambiguous figures, including Rorschach responses.

SUMMARY

An attempt was made to evaluate the usefulness of several theories offering explanations of color shock as it is said to occur in Rorschach testing. An arabic-greco-latin square design was utilized in the test series of a study

($N = 192$) which investigated the influence of color, pattern of color, figure, and anxiety level on response time. The influence on response time of stimulus attributes affecting complexity was confirmed. The hypotheses that highly anxious persons are greatly affected by colors are not confirmed. Partially verified were predictions made from Hullian behavior theory. It was found that drive level and figure difficulty were related to response time in that highly anxious persons became slower with increasing difficulty of figure during warm-up trials. During the test series, however, the predicted relationship was not found. Further work should explore more fully the extremes of the stimulus complexity and anxiety variables.

REFERENCES

1. ALLEN, R. M. The influence of color in the Rorschach test on reaction time in a normal population. *J. proj. Tech.*, 1951, 15, 481-485.
2. ALLEN, R. M., MANNE, S. H., & STIFF, M. The role of color on Rorschach's test: a preliminary normative report on a college student population. *J. proj. Tech.*, 1951, 15, 235-242.
3. BUKER, S. L., & WILLIAMS, M. Color as a determinant of responsiveness to Rorschach cards in schizophrenia. *J. consult. Psychol.*, 1951, 7, 196-202.
4. COX, F. N., & SARASON, S. B. Test anxiety and Rorschach performance. *J. abnorm. soc. Psychol.*, 1954, 49, 371-377.
5. DODGE, W. T. Influence of color upon form perception. Unpublished master's thesis, Univ. of Texas, 1953.
6. DUBROVNER, R. J., VON LACKUM, W. J., & JOST, H. A study of the effect of color in productivity and reaction time in the Rorschach test. *J. clin. Psychol.*, 1950, 6, 331-376.
7. FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, 45, 120-126.
8. GODBEER, E. Factors introducing conflict in the choice behavior of children. Unpublished master's thesis, Yale Univ., 1940.
9. HOLTZMAN, W. H. Emotional instability and perceptual blocking induced by color distraction. Paper read at South. Soc. Phil. Psychol., Austin, Texas, April, 1953.
10. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943. Pp. 185-225.
11. LAZARUS, R. S. An experimental analysis of the influence of color on the protocol of the Rorschach test. *J. Pers.*, 1948, 17, 132-135.
12. MEYER, B. T. An investigation of color shock in the Rorschach test. *J. clin. Psychol.*, 1951, 7, 367-370.
13. MIALE, FLORENCE R., & HARROWER-ERICKSON, MOLLY R. Personality structure in the psychoneuroses. *Rorschach Res. Exch.*, 1940, 4, 71-74.
14. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens, (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
15. NORMAN, R. D., & SCOTT, W. A. Color and affect: A review and semantic evaluation. *J. gen. Psychol.*, 1952, 46, 185-225.
16. PERLMAN, J. A. Color and validity of the Rorschach 8-9-10%. *J. consult. Psychol.*, 1951, 15, 122-126.
17. RABIN, A. I., & SANDERSON, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, 3, 216-225.
18. RORSCHACH, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.
19. SAPPENFIELD, B. R., & BUKER, S. L. Validity of the Rorschach 8-9-10% as an indication of responsiveness to color. *J. consult. Psychol.*, 1949, 13, 268-271.
20. SARASON, S. B., & GORDON, E. M. The test anxiety questionnaire: scoring norms. *J. abnorm. soc. Psychol.*, 1953, 48, 447-448.
21. SCHACHTEL, E. G. On color and affect: Contributions to an understanding of Rorschach's Test II. *Psychiatry*, 1943, 6, 393-407.
22. SHPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, 18, 358-382.
23. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
24. WALLEN, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, 43, 346-356.
25. WARNER, S. J. The color-preferences of psychiatric groups. *Psychol. Monogr.*, 1949, 63, No. 6 (Whole No. 301).

Received November 3, 1956.

AVOIDANCE CONDITIONING OF VERBAL BEHAVIOR WITHOUT AWARENESS: A PARADIGM OF REPRESSION¹

CHARLES W. ERIKSEN² AND JAMES L. KUETHE

The Johns Hopkins University

SEVERAL recent experiments have suggested that the principles of learning theory may be profitably applied to the study of language and verbal behavior (2, 4, 6). These studies have shown that the verbal behavior of a subject can be manipulated through the use of differential reinforcement and in many cases the subjects remain unaware that their behavior has been modified, i.e., they are unable to state in what way their behavior has changed or even that it has changed. While these experiments have dealt with overt verbal behavior, it seems reasonable that implicit verbal behavior or thinking might similarly be affected. If so, these experiments have some interesting implications for research on psychological defense mechanisms in that they offer suggestions as to how defenses might be learned and how they might operate.

It is an empirical observation that responses followed by punishment or anxiety tend to be eliminated or replaced by other responses. If thoughts or implicit verbal behavior are considered as responses subject to manipulation by reinforcement, we can expect that thoughts that have become conditioned stimuli for anxiety will be eliminated or replaced by other thoughts. Dollard and Miller (3) have presented a detailed theoretical treatment of defense mechanisms along these lines.

Such a theoretical account of the learning of defense mechanisms would gain support if it could be shown that thoughts or associations that were followed by anxiety were eliminated or replaced. It was our purpose in the present study to determine, first, whether verbal associations that the subject (*S*) had formed during his life experiences could be changed by punishment in the form of electric shocks, and second, whether these changes in verbal associations could be produced without the *S*'s becoming aware that his behavior had been modified. We also wished to determine if such

effects would persist when the experimental task was changed.

METHOD

The basic technique used in this study was a modified form of the word association test. Here the same set of stimulus words was administered to *S*s a number of times in an unbroken sequence. Table 1 shows the 15 stimulus words comprising the association list that was used. It was assumed that each word on this list constituted a stimulus having a hierarchy of associations or responses that had been acquired by *S* during his life experiences. From a pilot study we had found that when the same word list is repeated a number of times, *S*s tend to continue to repeat their original associations. We wished to determine whether strong electric shocks administered every time certain selected associations occurred would result in these responses dropping out and being replaced by other associations.

Prior to beginning the experimental sessions with an *S*, the *E* randomly designated five of the stimulus words in the list which were to be the *critical* stimuli for this particular *S*. During the first trial the *S*'s associations to these five words, whatever they happened to be, were punished by a strong electric shock. Without pause the *S* was given a second trial on the 15 stimulus words followed by a third trial and so on for a total of ten trials or until the *S* had reached a criterion of two successive trials without the occurrence of any one of the five punished associations. On each trial the stimulus words occurred in a different sequence. During these succeeding trials on the list, *S* was administered a strong electric shock every time he responded with one of the five associations that had been shocked on the first trial. Shocks were administered only when these "taboo" associations occurred.

Shocks were administered through ankle electrodes attached to the *S*'s left ankle. The source of the shock came from an inductorium and a 3 v.d.c. source. Faradic stimulation was used. The intensity of shocks varied slightly between *S*s due to such factors as differences in skin resistance, but in all cases it was evident from the *S*'s comments and behavior that the shocks were quite painful.

In order to maximize the opportunities for obtaining "learning without awareness," the *S*s were misinformed as to the nature of the experiment and the reason for the electric shocks. They were told that the purpose of the experiment was to determine the maximum speed of mental associations. They were instructed that shocks would occur under two conditions: one, if their association to a stimulus word was too slow for that particular stage of practice; and the second was a condition that the *E* could not divulge to them in advance, but they might discover it on their own during the session and thus be able to avoid getting the shocks.

¹ This research was sponsored by Research Grant M-1037 from the National Institute of Mental Health.

² Now at the University of Illinois.

TABLE 1
STIMULUS WORDS

Careless	Lovely
Perfect	Hasty
Cheerful	Gawky
Tidy	Rural
Rotten	Thrifty
Costly	Old
Small	Boastful
Polite	

The Ss were instructed to respond to each stimulus word with the first word that came to mind and as quickly as possible. To insure that they understood the nature of the task, several practice words were administered. They were also informed that the same list of words would be repeated over and over, but each time the words would occur in a different sequence. During the experimental session S's association time to each stimulus word was measured to the nearest hundredth of a second by means of an electronic voice key. The timer was started by the sound of E's voice and manually stopped by E when S responded.

At the conclusion of this learning phase of the experiment, Ss were informed that this experiment was over and were then asked to do one more task before leaving. This consisted of chained associations to the stimulus words. The Ss were told that the same stimulus words would be repeated again, but this time they were to say all the words that came to mind and continue doing so until the E said "stop." Fifteen seconds were allowed for chained associations to each stimulus word. Before starting this phase of the experiment, Ss were all carefully instructed that there would be no further electric shocks.

In carrying out the chained associations in this experiment, we were interested in two things. First, we wished to see if the punished association would occur when the S was required to give a large number of responses, and second, we were interested in determining if the effects of punishment would be generalized to a situation somewhat different from the original learning situation, in which the threat of shocks had been removed.

At the conclusion of the chained associations the S was closely interviewed to determine his ability to verbalize the true nature of the experiment and changes in his behavior. All Ss were asked the following questions: Did you discover how to avoid getting shocks? Did you do anything to keep from getting shocked? What did you do? Did you ever find yourself withholding a response word and trying to think of something else to say? (If a positive answer was obtained, S was asked if he could recall any instances where this occurred.) Those Ss who reported deliberately withholding responses and substituting other responses for them were also questioned as to whether they had to continue doing this throughout the experiment.

In addition to the questioning, Ss were shown a list of the stimulus words and asked to indicate which words had been followed by shock and what response word they had given when the shocks occurred.

Subjects. Thirty-one male undergraduate students

served as Ss in the experiment. All were obtained from the elementary psychology course at The Johns Hopkins University. As part of the course requirements, they were required to serve as Ss in psychological experiments.

RESULTS

Awareness. On the basis of the interview data it was possible to divide the Ss clearly into high- and low-level-of-awareness groups. The high awareness Ss were able to state the real basis for shocks and to describe what they had done to avoid them. They were also able to identify correctly three or more of the critical stimulus words as well as the response they had given that led to punishment. All reported deliberately withholding punished responses while trying to think of another response to substitute for it. However, they all reported that after a few trials it was no longer necessary to withhold the punished responses deliberately since the new, or safe, response began to occur "automatically."

The low awareness Ss, on the other hand, were unable to state any reasons for the shock other than the misinformation the E had given them prior to the experiment. Occasionally they expressed hypotheses that were quite unrelated to the real basis for shock and which were quite at variance with their actual behavior. They were unable to provide an accurate description of changes in their behavior and all stated that they had not learned how to avoid getting shocks. When asked to indicate which of the stimulus words had been followed by shock, their selection was random. In no case were their selections more than 20 per cent correct.

It was possible to classify eleven of the Ss as having a high level of awareness or insight and another eleven Ss as having a low level of insight. There were another five Ss who could not be clearly placed in either the insight or the noninsight group. These Ss seemed to have a vague idea as to the reasons for shock and what they needed to do to avoid them, but their verbalizations were not definite enough to permit clear-cut decisions. We have labeled these Ss as the partial insight group. The remaining four Ss resisted classification on our insight criteria. Three of these Ss had adopted an incorrect hypothesis concerning the shocks. While their hypotheses were incorrect, they were nonetheless effective in avoiding shocks. For example, two of these Ss had the hypoth-

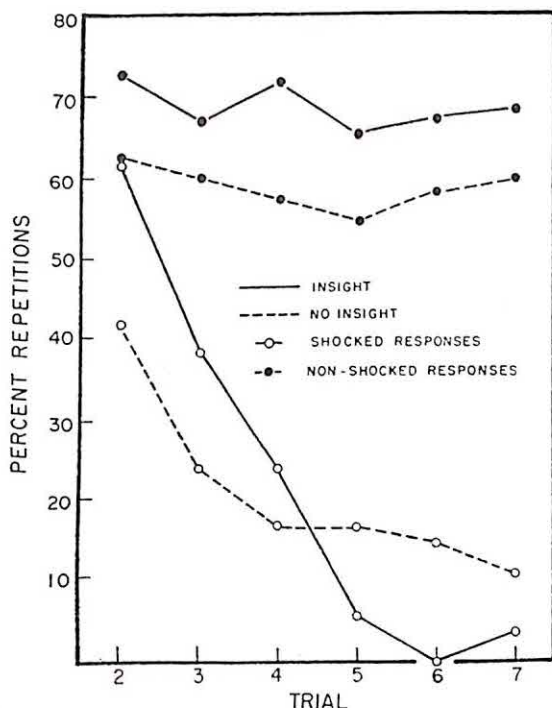


FIG. 1. PERCENTAGE OF FIRST-TRIAL RESPONSES REPEATED ON SUCCEEDING TRIALS AS A FUNCTION OF PUNISHMENT

esis that shocks occurred when they gave synonyms and they had changed all of their responses to avoid synonyms. The other two Ss had resorted to using only two response words, such as, "good" and "bad" for all stimulus words. The data from these four Ss were excluded from further analysis.

Since we were successful in obtaining two levels of awareness or insight in our Ss, we have treated insight as an independent variable in the remainder of the data analyses. We have not presented the data for the partial insight group, however, due to the small number of Ss that fell into this group.³

Avoidance learning. Both the insight and noninsight groups were quite successful in learning to avoid the punished associations. As Fig. 1 shows, the number of repetitions of punished responses decreases rapidly for both groups while the number of nonpunished responses that were repeated from trial to trial tends to remain constant. For the insight

group, a mean of 68.2 per cent of the first-trial, nonpunished response words were repeated on Trial 7 while only 3.6 per cent of the punished responses were repeated on this trial. A *t* test of the difference between these two means resulted in a *t* value of 12.2 which with ten degrees of freedom was significant beyond the .001 level. The same comparison for the noninsight group shows the mean percentage of repetition of the first-trial, nonpunished responses on Trial 7 to be 60.0 per cent as compared with a mean percentage of 10.9 for the punished responses. The *t* value of 6.21 is again quite significant ($p < .001$).⁴

While both the insight and noninsight groups show marked and reliable evidence of avoidance learning in this experimental situation, there does not appear to be any reliable evidence of a difference in rate of learning between these two groups. From Fig. 1, it appears that the noninsight group showed a tendency for more avoidance of shocked responses on Trials 2 through 4, but a *t* test based upon the mean percentage of shocked responses for each S in the two groups summed through these three trials does not approach statistical reliability ($t = 1.42$, $p > .10$). Also the same comparison between the insight and noninsight Ss on Trials 5 through 7 yields a *t* of 1.44 which is again insignificant ($p > .10$).

From Fig. 1 it appears that the noninsight group tended to give fewer first-trial responses to the noncritical stimuli than did the insight group. To determine whether this difference was reliable, the mean percentage of first-trial responses given by each S on Trials 2 through 7 was computed. The means of these mean percentages were 68.8 for the insight group and 58.8 for the noninsight group. The resulting *t* of 1.22 with 20 *df* does not approach

⁴ An over-all analysis of variance of these data was not justified due to the marked heterogeneity of variance that existed from the early to the late trials and the skewness of the distributions on the late trials. The *t* tests that have been employed have either been corrected for heterogeneity of variance by the method suggested by Cochran and Cox (1) or the comparisons are between data that show reasonable evidence of meeting the underlying assumptions involved.

As a further precaution all statistical analyses involving percentages in this study were repeated using an arc-sine transformation of the percentages. In no case did the analyses using the transformed data result in significance levels appreciably different from those reported for the untransformed data.

³ In nearly all of the subsequent comparisons it was found that the partial insight Ss were intermediate between the insight and the noninsight groups although their performance more closely resembled that of the insight Ss.

accepted levels of statistical reliability and we cannot conclude that the two groups differ on the repetition of first-trial associations on subsequent trials.

A further confirmation of the lack of effect of insight upon the rate of avoidance learning is found when we compare the number of trials required by the insight and the noninsight groups to reach the criterion of two successive trials without a repetition of a punished response. The results of this analysis are seen in Table 2. As this table shows, the mean number of trials required for both groups is nearly identical and the *t* test falls far short of acceptable levels of significance.

Reaction times. While there is no evidence of differences in rate of avoidance learning between the insight and the noninsight Ss, there are some marked differences between these two groups in reaction times to stimulus words that have had punished responses (critical stimuli). On each trial the median reaction time to critical and noncritical stimuli was computed for each S. In Fig. 2 the mean of these median reaction times as a function of number of trials is shown for the insight and the noninsight groups. For the noninsight group the mean reaction times for both the five critical and the ten noncritical stimuli decreases as a function of trials. By the last trial the reaction times are virtually the same for both critical and noncritical words.

For the insight group, however, the relation of reaction times to trials differs for the critical and the noncritical stimuli. The times for the noncritical words decrease with trials as did reaction times for both critical and noncritical words in the noninsight group, but for the critical stimuli there is an initial increase in reaction time. Only on the fourth trial does the mean reaction time to the critical words begin to decrease, and even then it is still longer than the reaction time on the first trial.

In order to evaluate the statistical reliability

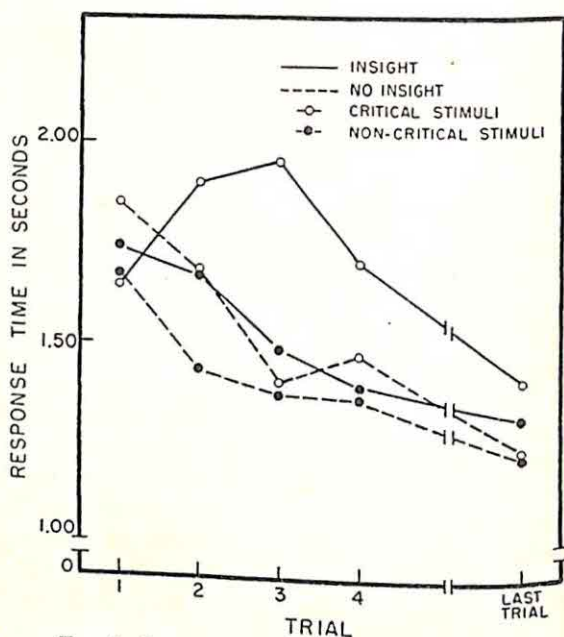


FIG. 2. REACTION TIMES TO CRITICAL AND NONCRITICAL STIMULI AS A FUNCTION OF TRIALS

of the interaction between insight, trials, and critical words that appears in Fig. 2, a modified four-way analysis of variance (insight, critical stimuli, trials, and Ss) was performed. The analysis used was identical to that described by Lindquist as Plan VI (5). The data for this analysis were the median reaction times for critical and noncritical stimuli for each S on each of the first four trials.⁵ The summary of this analysis is given in Table 3.⁶

Our main concern in this analysis was whether the trend of reaction times to the critical stimuli as a function of trials in the insight group was reliably different from the trend for the noncritical words and for both critical and noncritical words in the noninsight group. This question can be answered by the triple interaction between groups, critical stimuli, and trials. As is seen in Table 3, this interaction is highly significant ($p < .001$).

This difference in reaction time to critical and noncritical words between the insight and the noninsight groups provides us with an independent check upon the validity of classifying these Ss as insightful or insightfulless on the

TABLE 2

NUMBER OF TRIALS TO REACH THE CRITERION

Groups	Mean	Variance Estimate	<i>t</i>
Insight	4.18	2.8	
Noninsight	4.27	6.8	.09

Note.—The criterion trials have been omitted from the computations.

⁵ Only the first four trials were used in this analysis since a number of the Ss in both groups had reached the avoidance criterion by the end of the fourth trial.

⁶ Bartlett's test for heterogeneity of variance did not reach significance at the .05 level.

basis of their verbal statements. The increase in reaction time to critical stimuli in the insight group is what might be expected in view of their statements that they deliberately withheld responses that had been punished while they attempted to think of new responses. It should also be noted that the reaction-time data bear out their statements that they soon learned to think of nonpunished responses "automatically" since the reaction times for the last trials closely approach those for the noncritical stimuli.

The analysis in Table 3 also confirms several other effects that are indicated in Fig. 2. Trials are, of course, a significant source of variation, but there is no over-all difference between the insight and the noninsight groups other than the interactions with other variables. The main effect attributed to critical and noncritical words is significant at the .05 level, suggesting that the reaction times to the critical words tended to be somewhat longer during the first four trials. The only other significant effect is the interaction between trials and the insight and noninsight groups. Inspection of Fig. 2 shows that for the insight group, reaction times decreased at a slower rate for the first three trials than did the reaction times for the noninsight group.

TABLE 3

SUMMARY OF THE ANALYSIS OF VARIANCE OF THE REACTION-TIME DATA FOR TRIALS 1-4

Source	Sum of Squares	df	F Ratio	p
Total	47.32	175		
Trials	1.46	3	.49/.07 = 7.00	<.001
Critical stimuli	1.49	1	1.49/.21 = 7.09	<.05
Insight	1.03	1	1.03/1.36 = .76	>.10
Subjects	27.13	20	1.36/.07 = 19.42	<.001
Trials × critical stimuli	.29	3	.10/.07 = 1.43	>.10
Trials × insight	.95	3	.32/.07 = 4.57	<.01
Critical stimuli × insight	.07	1	.07/.21 = .33	>.10
Trials × critical stimuli × insight	2.06	3	.69/.07 = 9.86	<.001
Within-subject error terms:				
Pooled subjects by trials	4.25	60	.07/	
Pooled subjects by critical stimuli	4.15	20	.21/	
Pooled subjects by trials by critical stimuli	4.44	60	.07/	

TABLE 4

COMPARISON OF GROUPS ON THE AVERAGE NUMBER OF CHAINED ASSOCIATIONS PER STIMULUS AND THE AVERAGE PERCENTAGE OF FIRST-TRIAL RESPONSES

Groups	Mean Number of Associations		Mean Percentage of First-Trial Responses Repeated	
	Stimuli		Stimuli	
	Critical	Noncritical	Critical	Noncritical
Insight	4.5	4.4	49	66
Noninsight	4.4	4.5	50	74

Chained associations. Following the avoidance conditioning on the word-association test, the Ss were informed that there would be no more shocks in the experiment. They were then asked to chain-associate for 15 seconds to each of the stimulus words. In Table 4 we have presented a breakdown of the chained association data for the insight and noninsight Ss in terms of mean number of associations and mean number of first-trial response words repeated as a function of critical and noncritical stimuli.

The data on total associations and on percentage of repetitions of the first-trial words were separately analyzed by a modified three-way classification analysis of variance (insight, critical words, and Ss). The analysis for total associations revealed no significant effects. This is apparent from Table 4. However, the analysis of percentage of repetitions of first-trial responses resulted in one significant effect ($p < .01$). More first-trial responses to noncritical stimuli than to critical stimuli were elicited during chained associations. This finding indicates that the effects of shocks upon the Ss' associations have persisted despite a somewhat different task and the conscious knowledge that electric shocks will no longer occur.

DISCUSSION

In order to explain the phenomena of repression in terms of the principles used in describing avoidance conditioning and the effects of punishment as Dollard and Miller (3) have done, it is necessary to show that implicit verbal behavior or thoughts are analyzable into stimulus-response sequences. It is also necessary to show that these stimulus-response sequences are learned and modified by the same principles of reinforcement that

govern other behavior. In their book, Dollard and Miller have given a plausible treatment of these points.

The results of the present experiment have contributed another step toward an analysis of repression in terms of avoidance conditioning. There are two characteristics of repression that must be demonstrated by any experimental analogue of the repression process. In order for repression to be effective, the anxiety-provoking thought must be prevented from occurring. Thus, the experimental analogue must provide reasonable evidence that the changes in behavior have occurred not only at the overt level but at the covert level as well. Also, since a prominent feature of repression, as clinically observed, is the *S*'s lack of awareness of the repression process, the experimental analogue must demonstrate that the anxiety-provoking thought is eliminated "automatically." That is, the *S* should not have an awareness of having deliberately and intentionally changed his thoughts.

The results of the present experiment seem to meet these two requirements quite well, especially in the case of *S*s in the noninsight group. Not only did these *S*s show a marked and rapid decrease in punished associations, but this dramatic change in their associations occurred without verbal awareness on their part. These *S*s were unable under intensive questioning to describe what they had done to avoid shocks, and they reported no awareness of having deliberately withheld associations in an attempt to think of a new association.

Fortunately we do not have to rely upon the verbal reports of these *S*s alone to support our belief that this avoidance learning occurred without awareness. A comparison of the reaction times for the insight and the noninsight groups lends confirmation to the *S*s' verbal reports. The insight *S*s who had clear verbal awareness of how to avoid shocks and who reported having tried deliberately to withhold punished associations and think up new ones, were found to show an initial increase in reaction times to the critical stimuli for the first three trials. This increase in reaction time is what would be expected to accompany the implicit behavior that these *S*s describe having experienced. For the noninsight group, on the other hand, reaction times to the critical stimuli show a monotonic

decrease from the first trial to the last. The reaction times to the critical stimuli in this group are closely comparable to the reaction times to the noncritical stimuli throughout the experiment. The reaction-time data for the critical words in the noninsight group correspond quite well with what would be expected of an "automatic" repression process.⁷

While the verbalizations and the reaction-time data for the noninsight group strongly indicate that avoidance learning had occurred without awareness, these same data also suggest that the behavior changes occurred at the covert as well as at the overt level. Not only do the *S*s from the noninsight group report a lack of withholding of associations, but their reaction-time data also indicate that their overt and covert responses were identical. Had these *S*s implicitly responded with punished associations, it seems likely that their association times would have been appreciably longer to the critical stimuli than to the noncritical stimuli. Actually, however, there was little or no difference in the association times to the two classes of stimuli. Even among the *S*s in the insight group, the reaction times to critical stimuli had become comparable to the reaction times for noncritical stimuli by the last trial. This finding again corresponds with their verbal report that after a few trials the new associations "came to mind automatically" and they no longer thought of the punished responses.

One further piece of evidence that the shocks had effected associations on the implicit level is to be found in the data on the chained associations. Had these punished associations been present in implicit behavior, we would have expected more of them to have attained

⁷ It is interesting to equate the performance of the insight group with the process of conscious suppression and the performance of the noninsight group with the process of repression as these processes are described by Dollard and Miller (3). In both groups the punished responses were successfully avoided, but in the noninsight group the avoidance occurred "automatically" without awareness and without an increase in reaction times. In the insight group, on the other hand, the avoidance was a deliberate conscious act and was accompanied by increased reaction times while these *S*s searched for substitute responses. In the noninsight group the data indicate that the thought itself or the implicit verbal response was initially conditioned while in the insight *S*s the overt verbalization was first conditioned and then the thought or implicit response.

overt expression when the threat of punishment was removed. Both groups, however, gave fewer of the punished responses during chained association than would be expected on the basis of the number of first trial responses obtained to the noncritical stimuli.

While it was not the purpose of the present experiment to contribute to the theoretical controversy concerning the manner in which punishment eliminates responses, our data seem to favor an interpretation along the lines of an inhibitory process or decrement in the absolute response strength of the punished responses. An increase in response strength of alternative responses alone does not seem adequate to account for the data obtained from chained associations. Here the Ss had an opportunity to give a relatively large number of responses to each stimulus. Thus, it could be expected that not only would the stronger responses tend to occur, but there was the opportunity for responses lower in the response hierarchy to occur also. Yet only a few punished responses occurred, a result which is all the more surprising since these punished responses were for the most part initially the dominant response in the hierarchy.

SUMMARY

The Ss were presented with a 15-item word-association list under the guise that they were taking part in an experiment to determine the limit of speed of associations. They were instructed to respond with the first word that came to mind as quickly as possible after the stimulus word was presented. During the first trial on the list, the E administered a strong electric shock immediately after five arbitrarily selected response words. The Ss were then given a number of further trials on the same 15 stimulus words. Every time an S responded with one of the five first-trial punished responses, he received another electric shock. At the conclusion of this phase of the experiment, Ss were informed that there would

be no further shocks. They were then asked to chain-associate to each of the stimulus words.

On the basis of intensive questioning at the conclusion of the experiment, it was possible to classify the Ss into insight and noninsight groups. The insight Ss were characterized by having a high degree of verbal awareness as to the basis for the electric shocks and what they had to do to avoid receiving them. The noninsight Ss had a low level of verbal awareness as to the reasons for the shocks and what they had done to avoid them. Both insight and noninsight groups showed a rapid and marked learning of avoidance behavior. No reliable differences were found between the two groups in terms of rate of learning or on the number of trials to achieve the criterion of two successive trials without the occurrence of a punished response. However, there were clear differences between the two groups in reaction times to stimulus words that had elicited shocked responses. The nature of these differences corresponded with what would be expected from the Ss' verbalization. Both groups showed a significant decrement in the number of punished responses that occurred during chained associations. The relation of these data to a theory of repression was discussed.

REFERENCES

1. COCHRAN, W. G., & COX, GERTRUDE M. *Experimental designs*. New York: Wiley, 1950.
2. COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106-110.
3. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
5. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
6. SIDOWSKI, J. B. Influence of awareness of reinforcement upon verbal conditioning. *J. exp. Psychol.*, 1954, **48**, 355-360.

Received November 7, 1955.

THE EFFECTS OF VARYING COMBINATIONS OF AUTHORITARIAN AND EQUALITARIAN LEADERS AND FOLLOWERS¹

WILLIAM HAYTHORN
The RAND Corporation

ARTHUR COUCH
Harvard University

DON HAEFNER
PETER LANGHAM
University of Rochester
LAUNOR CARTER
The RAND Corporation

RECENTLY attention has been focused on the relationships between the personality characteristics of group members and the nature of the leadership that emerges from the interaction of these individuals with each other and their tasks (3, 8). The behavior of leaders in groups is determined not only by their own personalities, but also by the underlying needs, role expectations, and values of the group members. Conversely, the behavior of group members is partly determined by the personality of the leaders, particularly in interaction with their own needs and expectations. Of particular interest are groups in which there is some modal similarity or dissimilarity of the leader's and followers' personalities, creating situations where the interaction of different needs, expectations, perceptions, and values can be studied systematically. The present study was designed to investigate a specific case of the interaction between leader's and followers' personalities in the determination of behavior in groups.

This paper is the second report describing the results of a study of the behavior of authoritarian and equalitarian personalities in small groups. The California F scale was used as the principal measure of the authoritarian personality (1). In the first paper (11), behavioral differences between groups composed of individuals homogeneously high on the F scale and individuals homogeneously low on the F scale were reported. An analysis was made of the differences between these groups with respect to rated behavioral traits, recorded behavioral acts, and responses to a Post-Meeting Reaction Sheet. Significant differences between high-F and low-F Ss in

the predicted directions were found, indicating that a more "democratic" group culture prevailed in the low-F groups, and the low-F groups were more effective in dealing with the task problem. Significant differences—more marked than those distinguishing the groups as a whole—were found between the emergent leaders of the two kinds of groups, indicating that a qualitatively different kind of leadership occurs in groups composed of personalities at opposite ends of the authoritarian-equalitarian dimension. Emergent leaders in the low-F groups were more sensitive to others, more effective leaders, more prone to making suggestions for action subject to group sanction, and less likely to give direct orders to others.

To permit experimental manipulation of the leadership structure, these same subjects were subsequently assembled with appointed leaders in the study reported here. Half of the groups were composed with appointed leaders who had F scores similar to those of the other group members, while in the other half the appointed leaders had scores at the opposite end of the distribution. This design permitted an investigation of the effects of assembling groups with leaders whose personalities were similar to those of other group members, as compared with groups whose leaders' personalities were unlike those of other members. The major hypotheses examined were as follows:

1. *Hypothesis 1.* F-plus leaders (a) engage in less equalitarian behavior than F-minus leaders; (b) are less concerned with group approval for their actions; (c) engage in more autocratic behavior; and (d) show less sensitivity to others than do F-minus leaders.

2. *Hypothesis 2.* F-plus followers (a) engage in less equalitarian behavior; (b) show less sensitivity to others; (c) are more satisfied with appointed leaders; (d) are less critical of their own group; and (e) are more submis-

¹ The work described in this paper was done under a contract between the U. S. Navy, Office of Naval Research, and the University of Rochester. Dr. Launor F. Carter was the responsible investigator for that contract.

sive toward appointed leaders than F-minus followers.

3. *Hypothesis 3. Under F-minus leadership*, followers (a) have more influence on group processes; (b) develop a less formal group structure; and (c) express greater differences of opinion than under F-plus leadership.

4. *Hypothesis 4. With F-minus followers*, leaders (a) are not likely to engage in as much autocratic behavior; and (b) are not as distinct from other group members as are leaders with F-plus followers. That is, leaders with F-plus followers are thought, when under pressure, to take decisive, directive action in order to get the job done, because of the expectations on the part of F-plus followers regarding the role of an appointed leader.

5. *Hypothesis 5. In groups composed of leaders and followers with similar F-Scale scores*, (a) followers are more secure; (b) followers are more motivated to achieve the common group goal; (c) followers are more satisfied with the appointed leadership; (d) there is less personality conflict within the group; (e) there is less conflict between the leader and group members; and (f) there is higher group morale and productivity than in groups composed of leaders and followers with dissimilar F-scale scores.

METHOD

Subjects. The subjects (Ss) for the experiments consisted of (a) 32 male undergraduate students who were high on the F scale, "conservative" as judged by the Cattell Q₁ scale, and relatively normal as estimated by the MMPI; and (b) 32 male undergraduate students who had low F-scale scores, were liberal as determined by the Q₁ scale, and were also "normal" as judged by the MMPI. The mean F-scale scores for the two groups were 124.5 and 66.9, respectively, using the 30-item, 7-points-per-item scale described in the *Authoritarian Personality* (1).²

The task. The task given the groups was the discussion and composition of a short script involving a human relations problem. Complete instructions were read to the Ss before they began working, to the effect that they were to help E in developing a test of human relations skills. They were to be shown a film of a human relations problem, after which they were to answer a number of questions about how the problem had been handled, how it should have been handled, etc. When they finished answering the questions as individuals, they were to discuss similar questions as a group and arrive at answers to which they all agreed. Following completion of this phase of the task, they were to com-

pose dialogue for similar films. Once their script was completed, they were to record it on a magnetic tape recorder. After reading the instructions, E showed the film, then left the room. He did not return until the group had completed all phases of the procedure outlined in the instructions. A typed copy of the latter was left on the table with the Ss.

Observation procedures. While Ss worked on the task in the experimental room, observers watched through one-way vision mirrors from an adjacent room. Four Os (first- and second-year graduate students) recorded their observations in three ways, as follows:

1. While observing, Os categorized and "typed out" behavioral units as they occurred, using a Stenotype system of interaction recording described in an earlier paper (4). The category system was revised, however, to fit the needs of this study.

2. At the end of 30 minutes of recording, the pair of Os was relieved by another pair. Immediately after being relieved, Os rated each S on 16 behavioral characteristics, defined in the previous report (11), as follows: striving for individual prominence, friendliness, security, influence, equalitarianism, striving for group approval, submissiveness, striving for goal achievement, self-isolating behavior, effective intelligence, sensitivity to others, leadership, aggressiveness, autocratic behavior, social ability, and nonadaptability. These ratings had an average interobserver reliability of .75, with a range from .31 to .91. Since the group sessions varied in length from two to four hours, each S was rated by each O at least once in each group, and usually two or more times.

3. At the end of each group session, all four Os and E filled out a Post-Meeting Reaction Sheet (PMRS) designed to describe aspects of the group not easily attributed to single individuals. Examples of the items on the PMRS are:

a. This group was not serious enough—there was too much playing around.

b. The group effectiveness was interfered with by personality clashes between members.

There were 37 such items, to each of which Os indicated the extent of their agreement, using a 7-point scale.

4. In addition, Ss were asked to fill out a very similar PMRS, and to rate the degree to which they were satisfied with their group, the extent to which they thought the group had been productive, to rank the group members in order of how well they were liked, and to rank group members in terms of their relative contributions to the achievement of group goals.

"Emergent" group sessions. The Ss were scheduled to come to the laboratory work room in groups of four. In their first session—results of which are described in the previous paper (11)—each of the sixteen groups were composed of Ss homogeneous with regard to F-scale scores. In this session, there was no experimental structuring of leadership.

Appointed leader sessions. After the emergent sessions, all individuals were reassembled with three members from their original emergent groups and another member from a different group. The alien member was appointed as leader (by telling him he was to take charge of group activities that day and would be responsible for getting the job done). In half of the groups

² Further details regarding the selection of subjects are given in Reference 11.

the leader had an F score similar to that of group members, and in half it was dissimilar. Thus there four groups with F-plus leaders and F-plus followers, four with F-plus leaders and F-minus followers, four with F-minus leaders and F-plus followers, and four with F-minus leaders and F-minus followers.

In Session III, two of the original emergent-group members who had been together in Session II were reassembled with the member who had been withdrawn during Session II. In addition, an alien member was introduced as the leader for that session. Groups having had an F-plus leader in Session II were given an F-minus leader in Session III, and vice versa.

Characteristics of appointed leaders. The experimental design required that 32 of the 64 Ss be appointed as leaders, 16 in Session II and 16 in Session III. Two were chosen from each of the original groups as follows: The men in each group were assigned ranks on the basis of the leadership nominations made by Ss in their PMRS, and also on the basis of leadership ratings made by Os. Finally, in the F-minus groups each S was ranked on the basis of the equalitarian ratings given by Os, while in the F-plus groups they were ranked on the basis of ratings by Os of autocratic behavior. These three sets of ranks were averaged, and the two individuals (Ranks I and II) with the lowest average ranks were selected to serve as appointed leaders in subsequent groups. A later check showed that this procedure, designed to assure as nearly as possible that the appointed leaders would behave in an authoritarian or equalitarian manner, resulted in selecting the same leaders as would have been chosen by subject ratings alone if ties were broken by observer ratings.

Analysis of data obtained in the emergent sessions (11) indicated significant behavioral differences between the individuals chosen from the F-plus groups and those chosen from the F-minus. In summary, leaders chosen from the F-minus groups were rated as being more friendly, more secure, more equalitarian, less striving for group approval, more submissive, less self-isolating, showing more effective intelligence, showing greater sensitivity to others, and showing greater leadership. The pattern of results generally agrees with current opinion concerning the differences between authoritarian and equalitarian leaders, except that the differences in aggressiveness and autocratic behavior failed to reach statistical significance.

Observers were not informed of the composition of the groups during the experiment, and were unable to guess the composition pattern significantly more accurately than chance. This finding may be attributed to an observer expectation that authoritarian individuals would be more aggressive, argumentative, and generally uncompromising towards others. Instead, in the emergent sessions they were rated as being significantly more concerned with group approval than were the F-minus Ss. Such concern for group approval in groups having no formal authority structure is theoretically a central facet of the authoritarian personality structure (1), underlying the tendency to conform to social mores. However, our Os apparently tended to identify pleasant social behavior as evidence of equalitarianism and consequently made many errors in attempting to estimate group composition. The fact that analyses generally supported our hypotheses in

spite of observer failure to distinguish the groups is presumptive evidence that results to be reported are not attributable to bias.

RESULTS

Trait ratings of appointed leaders. The experimental design, a 2×2 factorial design with leaders' and followers' personalities as the main effects, permitted testing of several sources of variance. Group means and results of the analyses of variance of ratings by Os are presented in Table 1. Only the ratings yielding significant *F* ratios are included, in order to conserve space. Seven of the 16 rating variables did not yield significant *F* ratios for any of the three sources on either session.

The data entered in Table 1 are mean ratings of the appointed leaders, averaged over four Os. The four groups are F-plus leaders with F-minus followers, etc. The within-group variance is taken as the best estimate of error, or variance not attributable to experimental conditions. The "effects on leaders" are divided into three sources: *F* scores of the followers, *F* scores of the leaders, and whether the leader and followers were homogeneous or heterogeneous with regard to *F* scores (the latter is the leader-by-follower interaction in the analysis of variance).

Trait ratings of followers. Analysis of the ratings of followers' behavior by Os parallels that for ratings of leaders. The main effects are the *F* scores of the followers and of the leaders. Again the interaction mean square indicates the effects of homogeneity vs. heterogeneity of leader-follower combinations. Only seven of the 16 variables yielded significant *F* ratios. These are presented in Table 2, with the group means and the within-group variance estimates.

There were three followers in each of the 16 groups in each session. Thus, each of the means of the four conditions are based on ratings of 12 individuals. However, in an earlier study (9) it was shown that the ratings of individuals within a group are not independent. Therefore, the average rating of the three followers in each group was computed and these averages used in computing means and within-group variance. Each condition mean is based on the averages of four groups. This procedure results in 15 degrees of freedom ($16 - 1$), distributed as follows: 1 for leaders' *F* scores, 1 for followers' *F* scores, 1 for interac-

TABLE 1
TRAIT-RATING RESULTS: APPOINTED LEADERSHIP SESSIONS
(Leaders)

Analysis of Variance		Leader Means					Effect on Leaders of:		
Trait names	Session	F+ leaders		F- leaders		s ² w	F+ vs. F- followers F ratio	Leader's personality F ratio	Group's composition F ratio
		(F + F)†	(F- F)†	(F + F)	(F- F)				
I. Individual Prominence	II	4.76	3.19	4.08	4.13	0.16	14.574(F+)**	0.028(F-)	1.097(Homo)
	III	4.51	3.08	4.28	3.52	0.77	6.298(F+)*	0.056(F-)	0.572(Homo)
V. Equalitarianism	II	3.56	4.76	5.29	4.82	0.70	0.747(F-)	4.569(F-)	3.954(Hetero)
	III	3.27	4.04	4.89	5.48	1.00	1.837(F-)	9.350(F-)*	0.033(Hetero)
VI. Striving for Group Approval	II	3.89	3.62	4.49	4.26	0.25	1.039(F+)	6.264(F-)*	0.008(Homo)
	III	4.12	3.31	4.42	4.24	0.72	1.360(F+)	2.068(F-)	0.567(Homo)
VII. Submissiveness	II	3.03	3.28	3.43	2.75	0.59	0.325(F+)	0.026(F+)	1.462(Hetero)
	III	2.57	3.77	3.34	3.07	0.41	2.100(F-)	0.011(F-)	5.165(Hetero)*
IX. Aggressiveness	II	2.98	1.78	1.92	2.49	0.39	1.013(F+)	0.329(F+)	8.037(Homo)*
	III	2.95	1.63	1.96	1.34	0.51	7.441(F+)*	3.223(F+)	0.964(Homo)
X. Autocratic	II	3.64	1.99	2.21	2.61	0.56	2.773(F+)	1.150(F+)	7.459(Homo)*
	III	3.56	1.98	2.20	1.64	0.43	10.821(F+)*	6.877(F+)*	2.471(Homo)
XI. Effective Intelligence	II	4.32	4.46	5.23	4.92	0.31	0.095(F+)	5.963(F-)*	0.653(Hetero)
	III	4.21	3.90	4.96	5.42	0.53	0.038(F-)	9.759(F-)**	1.099(Homo)
XII. Sensitivity	II	3.27	4.26	4.89	4.47	0.77	0.423(F-)	4.361(F-)	2.607(Hetero)
	III	3.28	3.85	4.80	5.39	0.83	1.616(F-)	11.304(F-)**	0.001(Homo)
XV. Nonadaptability	II	2.01	1.38	1.31	1.70	0.14	0.407(F+)	1.055(F+)	7.469(Homo)*
	III	1.45	1.75	1.45	1.08	0.60	0.008(F+)	0.759(F+)	0.759(Hetero)

* p (.05) = 4.75: df 1/2.

** p (.01) = 9.33.

† (F + F), (F - F), refers to the personality of Followers, (F+ Followers) (F- Followers), respectively.

tion, and 12 for within-group variance. The estimate of significance so derived is believed to be very conservative.

Behavioral category indices for followers. It will be recalled that Os categorized and recorded behavioral acts as they occurred. Forty-three categories were used, but in order to simplify the analysis and increase the frequency of acts per category, these were combined into 17 category indices: I. Friendly acts; II. Positive affect acts; III. Direction-taking acts; IV. Self-emphasizing acts; V. Asking for suggestion or sanction; VI. Asking for group evaluation; VII. "Democratic" acts; VIII. Directive acts; IX. Expressions of confusion or lack of orientation; X. Hostile acts; XI. Negative affect acts; XII. Initiating and integrating acts; XIII. Diagnosing and clarifying acts; XIV. General task participation acts; XV. Self-isolating acts; XVI. Withdrawing and out-of-field activity; and XVII. Tension release and out-of-field activity.

ity. These indices are not independent, in that some categories of behavior were included in more than one index. For a more complete description of the categories and indices, the reader is invited to consult the earlier report (11). Results of the analysis of these indices are presented in Table 3.

Subject's Post-Meeting Reaction. At the end of each session, all four Ss were asked to complete a Post-Meeting Reaction questionnaire, asking them to express the degree to which they agreed to a number of statements about the group. Agreement was expressed on a 7-point scale with the middle point missing to eliminate noncommittal responses (except the items on motivation and productivity, which had 9 points). Items on the questionnaire were grouped on a priori grounds into indices, and the means of the three followers in each group were computed. These follower means were then analyzed in the same way as previous data. Results of this analysis for all

TABLE 2
TRAIT-RATING RESULTS: APPOINTED LEADERSHIP SESSIONS
(Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Trait names	Session	F+ followers		F- followers		s ² w	F+ vs. F- leadership <i>F</i> ratio	Follower's personality <i>F</i> ratio	Group's composition <i>F</i> ratio
		(F+ L)†	(F- L)†	(F+ L)	(F- L)				
III. Security	II	3.90	4.54	4.10	4.25	0.17	3.678(F-)	0.046(F+)	1.428(Hetero)
	III	4.54	4.30	4.17	4.78	0.07	2.146(F-)	0.196(F-)	10.484(Homo)**
IV. Influence	II	3.85	3.91	3.92	3.81	0.09	0.028(F+)	0.014(F+)	0.310(Hetero)
	III	4.03	4.11	3.73	4.28	0.06	6.544(F-)*	0.262(F+)	3.671(Homo)
V. Equalitarianism	II	2.62	3.16	3.57	3.52	0.45	0.525(F-)	3.831(F-)	0.777(Hetero)
	III	2.64	2.61	2.94	3.99	0.32	3.211(F-)	8.711(F-)*	3.600(Homo)
VI. Striving for Group Approval	II	3.12	3.16	2.72	3.04	0.16	0.823(F-)	1.717(F+)	0.480(Homo)
	III	3.50	3.30	2.94	3.45	0.08	1.241(F-)	2.046(F+)	6.165(Homo)*
VIII. Striving for Goal Achievement	II	3.34	3.63	3.94	3.90	0.27	0.218(F-)	2.827(F-)	0.398(Hetero)
	III	3.90	3.84	3.90	4.64	0.13	3.628(F-)	4.881(F-)*	4.881(Homo)*
XI. Effective Intelligence	II	3.25	3.48	3.97	3.62	0.20	0.072(F+)	3.636(F-)	1.702(Hetero)
	III	3.42	3.44	3.47	4.25	0.11	5.589(F-)*	6.537(F-)*	5.244(Homo)*
XII. Sensitivity	II	2.83	3.07	3.67	3.36	0.42	0.001(F+)	3.093(F-)	0.726(Hetero)
	III	2.92	2.96	3.30	4.08	0.31	2.184(F-)	7.338(F-)*	1.780(Homo)

* p (.05) = 4.75: df 1/2.

** p (.01) = 9.33.

† (F + L), (F - L), refers to personality of leaders: (F + Leaders) (F - Leaders), respectively.

TABLE 3
BEHAVIOR CATEGORY INDICES: APPOINTED LEADERSHIP SESSIONS
(Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Index names	Session	F+ followers		F- followers		s ² w	F+ vs F- leadership <i>F</i> ratio	F+ vs. F- personality <i>F</i> ratio	Group's composition <i>F</i> ratio
		(F+ L)	(F- L)	(F+ L)	(F- L)				
VI. Asking for Group Evaluation (14, 11)	II	1.28	1.86	2.83	2.58	0.66	0.16(F-)	7.87(F-)*	1.07(Hetero)
	III	2.87	1.73	1.93	2.60	3.18	0.07(F+)	0.00	1.03(Homo)
VII. "Democratic" Acts (34, 12, 14, 11)	II	2.24	2.67	3.71	3.68	1.11	0.14(F-)	5.54(F-)*	0.20(Hetero)
	III	5.00	2.78	3.22	3.37	6.85	0.62(F+)	0.20(F+)	0.82(Homo)
IX. Expressions of Confusion or Lack of Orientation (22, 33, 61)	II	1.43	2.28	1.58	1.40	0.56	0.81(F-)	0.96(F+)	1.91(Hetero)
	III	2.75	0.89	1.29	1.97	1.01	1.38(F+)	0.14(F+)	6.35(Homo)*
XIII. Diagnosing and Clarifying Acts (35, 38)	II	2.21	3.12	5.90	4.38	2.32	0.16(F+)	10.60(F-)**	2.56(Homo)*
	III	2.47	1.57	3.04	3.87	3.76	0.00	2.18(F-)	0.80(Homo)
XVI. Withdrawing and Out-of-field Activity (75, 76)	II	20.21	7.09	5.01	8.06	45.06	1.87(F+)	5.08(F+)*	5.18(Homo)*
	III	8.93	12.74	7.31	14.97	79.93	1.65(F-)	0.00	0.19(Homo)
XVII. Tension Release and Out-of-field Activity (8, 76)	II	51.74	21.24	11.45	11.92	258.73	3.49(F+)	9.51(F+)*	3.70(Homo)
	III	32.14	39.39	22.68	28.56	276.56	0.62(F-)	1.49(F+)	0.01(Hetero)

* p (.05) = 4.75: df 1/2.

** p (.01) = 9.33.

indices yielding significant F ratios are given in Table 4. No significant results were obtained for 12 of the 21 indices from the Ss' PMRQ. These 12 were: II. Degree of equal participation; III. Degree of personality conflict; IV. Degree of informal friendliness; VII. Satisfaction with leadership; VIII. Degree of conflict within group; IX. Competence of members; X. Morale; XV. Formation of cliques; XVI. Personality conflict between leader and group members; XXI. Differences of opinion; XXII. Competition among members; and XXIII. Lack of cooperation.

Observers' Post-Meeting Reactions. Four Os also completed the PMRS. The same indices were computed, with the exception that Os answered questions regarding the formality of group structure and communication effectiveness, and did not answer the question concerning relative efficiency of leadership. Results of the Os' PMRS analysis are presented in Table 5. Nineteen of the 23 indices

yielded at least one significant F ratio. Those that did not were: XVIII. Degree of goal-oriented control by appointed leader; XIX. Domination by appointed leader; and XXII. Competition among members.

Experimental findings. The results, in general, support the hypotheses stated at the beginning of the paper. With regard to Hypothesis 1, F-plus leaders were rated as being less equalitarian, less concerned with group approval, more autocratic, and less sensitive to others than F-minus leaders (Table 1). They were also rated as showing less effective intelligence (Table 1). The latter result is believed to be partially a function of the particular task given the groups, in this case a task requiring analysis of basic values which is presumably alien to F-plus personalities.

F-plus followers, as predicted from Hypothesis 2, were rated as being less equalitarian behaviorally, and as showing less sensitivity to others (Table 2). They were also rated as

TABLE 4
SUBJECTS' POST-MEETING REACTION SHEET INDICES: APPOINTED LEADERSHIP SESSIONS
(Followers)

Analysis of Variance		Group Means					Effect on Followers of:		
Index names	Session	F+ followers		F- followers		s ² w	F+ vs. F- leadership <i>F</i> ratio	Follower's personality <i>F</i> ratio	Group's composition <i>F</i> ratio
		(F+ L)†	(F- L)†	(F+ L)	(F- L)				
I. Dissatisfaction with Goal Progress	II	3.13	2.30	2.52	2.92	0.23	0.81(F+)	0.00	6.53(Homo)*
	III	2.53	3.28	2.50	2.38	0.25	1.59(F-)	3.45(F+)	2.97(Hetero)
V. Definiteness of Leadership	II	5.17	5.35	3.98	4.00	0.10	0.42(F-)	62.18(F+)**	0.27(Hetero)
	III	4.40	4.31	4.33	3.64	0.10	6.03(F+)*	5.40(F+)*	3.70(Hetero)
VI. Striving for Equal Participation	II	5.96	5.75	4.42	5.21	0.36	0.95(F-)	12.06(F+)**	2.77(Homo)
	III	5.96	5.79	4.54	5.42	0.65	0.77(F-)	4.93(F+)*	1.67(Homo)
XI. Group Productivity	II	6.83	7.83	6.75	5.92	0.55	0.05(F-)	7.26(F+)*	6.10(Hetero)*
	III	7.25	6.67	6.92	6.75	1.15	0.49(F+)	0.05(F+)	0.15(Homo)
XIV. Motivation Towards Group Goal	II	5.50	5.17	3.83	4.00	1.62	0.02(F+)	4.96(F+)*	0.15(Homo)
	III	5.58	5.33	4.25	4.58	1.11	0.01(F-)	3.90(F+)	0.31(Homo)
XVII. Relative Efficiency of Leadership	II	5.50	5.50	6.08	4.25	0.68	4.94(F+)*	0.65(F+)	4.94(Hetero)*
	III	5.41	3.55	4.16	5.41	0.58	0.60(F+)	0.60(F-)	15.46(Homo)**
XVIII. Degree of Goal-oriented Control by Appointed Leader	II	4.83	6.04	5.44	4.94	0.39	1.28(F-)	0.64(F+)	7.43(Hetero)*
	III	5.56	4.41	4.69	5.00	0.48	1.34(F+)	0.16(F+)	4.10(Homo)
XIX. Domination by Appointed Leader	II	3.21	3.46	2.29	2.67	0.17	2.27(F-)	16.97(F+)**	0.09(Homo)
	III	2.79	2.72	2.42	2.04	0.68	0.27(F+)	1.50(F+)	0.12(Hetero)
XX. Satisfaction with Appointed Leader	II	4.58	5.67	4.50	3.50	0.78	0.01(F-)	6.45(F+)*	5.53(Hetero)*
	III	5.25	5.22	3.91	3.66	1.34	0.05(F+)	5.79(F+)*	0.03(Hetero)

* p (.05) F ($\frac{1}{2}$) = 4.75.

** p (.01) = 9.33.

† (F + L), (F - L), refers to personality of leaders (F + leaders), (F - leaders), respectively.

TABLE 5
OBSERVERS' POST-MEETING REACTION SHEET INDICES: APPOINTED LEADERSHIP SESSIONS
(Groups)

Analysis of Variance		Group Means					Effect on "Groups" of:		
Index names	Session	F+ "Groups"		F- "Groups"		s ² w	Followers		Group's composition F ratio
		(F+ L)	(F- L)	(F+ L)	(F- L)		F+ vs. F- leadership F ratio	F+ vs. F- personality F ratio	
I. Dissatisfaction with Goal Progress	II	3.85	2.48	2.52	2.92	0.76	1.23(F+)	1.04(F+)	4.10(Homo)
	III	3.23	3.81	2.21	2.21	0.81	0.42(F-)	8.50(F+)*	0.42(Hetero)
II. Degree of Equal Participation	II	2.88	3.02	3.17	3.08	1.20	0.01(F-)	0.10(F-)	0.04(Hetero)
	III	3.38	3.48	2.28	3.78	0.33	7.71(F-)*	1.93(D+)	5.90(Homo)*
III. Degree of Personality Conflict	II	3.73	3.48	2.90	3.58	0.64	0.29(F-)	0.83(F+)	1.35(Homo)
	III	3.21	4.22	3.54	2.54	0.20	0.00	9.29(F+)*	20.59(Hetero)*
IV. Degree of Informal Friendliness	II	4.73	4.62	4.90	4.40	0.68	0.56(F+)	0.00	0.22(Hetero)
	III	5.00	4.47	4.32	5.48	0.36	1.11(F-)	0.31(F-)	7.98(Homo)*
V. Definiteness of Leadership	II	5.52	5.14	4.04	5.06	0.40	1.00(F-)	6.04(F+)*	4.93(Homo)*
	III	5.09	4.84	4.58	4.60	0.42	0.12(F+)	1.34(F+)	0.18(Homo)
VI. Striving for Equal Participation	II	3.40	4.85	2.88	3.53	0.75	7.81(F-)*	2.51(F-)	0.30(Hetero)
	III	3.98	4.72	4.32	5.78	0.89	5.43(F-)*	2.20(F-)	0.55(Homo)
VII. Satisfaction with Leadership	II	4.95	5.04	5.60	4.85	0.68	0.13(F+)	0.01(F-)	2.58(Homo)
	III	5.70	4.35	5.10	5.80	0.26	1.61(F+)	2.75(F-)	16.01(Homo)**
VIII. Degree of Conflict Within Group	II	3.26	3.29	2.69	4.09	0.73	2.77(F-)	0.07(F-)	2.58(Homo)
	III	2.81	3.76	3.18	2.38	0.40	0.06(F-)	2.64(F+)	7.71(Hetero)*
IX. Competence of Members	II	4.70	5.85	5.58	5.15	0.34	1.50(F-)	0.10(F-)	7.32(Hetero)*
	III	4.93	4.60	5.60	6.18	0.37	0.17(F-)	13.85(F-)*	2.30(Homo)
X. Morale	II	4.30	4.70	4.48	3.92	0.84	0.03(F+)	0.43(F+)	1.12(Hetero)
	III	4.80	4.00	4.45	5.72	0.21	1.06(F-)	9.04(F-)*	20.68(Homo)**
XI. Group Productivity	II	6.15	7.35	6.25	5.50	1.60	0.16(F-)	1.81(F+)	2.26(Hetero)
	III	6.20	5.15	6.50	7.90	0.78	0.16(F-)	11.99(F-)**	7.74(Homo)*
XII. Formality of Group Structure	II	3.10	2.25	2.50	2.40	0.44	0.97(F+)	1.29(F+)	0.46(Homo)
	III	2.85	2.30	2.90	2.00	0.34	6.26(F+)*	0.19(F+)	7.69(Homo)*
XIII. Communication Effectiveness	II	5.40	5.40	5.50	4.90	0.49	0.74(F+)	0.33(F+)	0.74(Hetero)
	III	5.40	4.95	4.75	5.95	0.35	1.59(F-)	0.35(F-)	7.69(Homo)*
XIV. Motivation Towards Group Goal	II	3.65	5.15	4.90	5.05	0.90	3.03(F-)	1.47(F-)	2.03(Hetero)
	III	3.80	4.00	4.55	5.95	0.44	5.75(F-)*	16.38(F-)**	3.24(Homo)
XV. Formation of Cliques	II	2.90	3.60	2.90	3.45	1.32	1.18(F-)	0.02(F+)	0.02(Hetero)
	III	1.90	3.20	2.30	2.10	0.30	3.99(F-)	1.62(F+)	7.42(Hetero)*
XVI. Personality Conflict Between Leader and Group Members	II	2.72	3.15	2.33	3.03	0.84	1.52(F-)	0.30(F+)	0.08(Homo)
	III	2.12	3.83	3.05	2.15	0.23	2.86(F-)	2.41(F+)	29.38(Hetero)**
XX. Satisfaction with Appointed Leader	II	3.15	2.25	3.55	4.35	1.06	4.37(F-)	0.06(F-)	0.29(Homo)
	III	3.35	4.00	3.50	5.65	1.52	5.15(F-)*	2.13(F-)	1.48(Homo)
XXI. Difference of Opinion	II	3.15	3.70	2.60	4.65	0.87	7.76(F-)*	0.18(F-)	2.58(Homo)
	III	2.75	3.62	3.22	2.52	0.43	0.07(F-)	0.92(F+)	5.82(Hetero)*
XXIII. Lack of Cooperation	II	4.05	3.25	3.10	3.75	1.19	0.02(F+)	0.17(F+)	1.77(Homo)
	III	3.25	4.30	3.70	2.30	0.41	0.30(F+)	5.82(F+)*	14.55(Hetero)**

* $p(.05) F(1/2) = 4.75$.** $p(.01) = 9.33$.

showing less goal motivation (Table 2). Again, the latter results were probably highly contingent on the task. The category indices supported these results by indicating that F-plus followers engaged in less asking for group evaluation and "democratic" acts than did F-minus followers (Table 3).

F-plus followers were also found to be more satisfied with appointed leaders than F-minus followers, and were apparently less critical of their own groups in the sense that they rated their groups as more productive and more motivated to achieve the group goal than did the F-minus followers (Table 4). That this result is due to differential response tendencies is suggested by the fact that Os rated the groups with F-minus followers as showing more group productivity and goal striving (Table 5), and recorded more withdrawing and out-of-field activity for F-plus followers (Table 3).

Hypothesis 2e, that F-plus followers would be more submissive toward appointed leaders, was not supported by Os' ratings of submissiveness nor by the recorded "direction-taking" behaviors. There was some support for this hypothesis from the PMRS analyses, however, where F-plus followers rated their groups as having more definite leadership and as being more dominated by the appointed leader (Table 4), and Os rated groups with F-plus followers as having more definite leadership (Table 5).

It was hypothesized (3a) that under F-minus leaders, followers—whether they were F-plus or F-minus—would exert more influence on group processes. This prediction was supported by ratings of "influence" and "effective intelligence" (Table 2). The hypothesis is further supported by the finding that Os rated groups with F-minus leaders as having a significantly higher degree of equal participation (Table 5).

Hypothesis 3b—that groups with F-minus leaders would develop a less formal group structure—was supported by the results with Os' PMRS rating (Table 5). By "formal group structure" Os attempted to describe the extent to which the leader played his task role in a stilted manner. This finding does not imply any differences in the level or definiteness of leadership. Rather, it seems

that the crucial differences here are in the style in which authoritarian and equalitarian personalities display leadership.

Hypothesis 3c, that followers would express greater differences of opinion under F-minus than under F-plus leaders, was supported by Os' ratings on the PMRS (Table 5), but not by the Ss' PMRS ratings and not by the behavioral category index of "negative affect acts," which subsumes the category "disagrees." The category index results were in the predicted direction, but did not approach statistical significance. Hypothesis 3c is, therefore, considered of doubtful tenability.

Hypothesis 4 stated that leaders with F-minus followers would differ from those with F-plus followers by (a) engaging in less autocratic behavior, and (b) being less distinct from other group members. These predictions were both supported by findings that leaders with F-plus followers were rated higher on autocratic behavior (Table 1), that Ss and Os both rated groups with F-plus followers as having more definite leadership (Tables 4 and 5), and that F-plus followers rated their groups as being more dominated by the appointed leader (Table 4). (The latter results were offered as support for Hypothesis 2e above. Their repetition here is indicative of the interactive nature of the hypothesis investigated.)

Hypothesis 5a—that followers are more secure in homogeneous groups—was supported by the difference in rating of "security" by Os (Table 2). Followers in homogeneously composed groups were also rated as "striving for goal achievement" more than those in heterogeneous groups (Table 2), in support of Hypothesis 5b.

The hypothesis that followers would be more satisfied with the appointed leaders in homogeneous groups was not supported. In fact, quite the opposite was found, i.e., followers in the heterogeneous groups rated themselves as significantly more satisfied with their appointed leader. No satisfactory explanation of this result has been achieved, but inspection of the means indicates that the largest differences in satisfaction were between F-plus followers with F-minus leaders and F-minus followers with F-minus leaders. The latter were relatively quite dissatisfied with

their leaders. To what extent this finding is due to differential response tendencies cannot be determined with the data obtained.

Hypothesis 5*d*—that there is less personality conflict in homogeneous groups—was supported by Os' PMRS ratings (Table 5), but not by Ss' ratings. Os' ratings are believed to provide better tests of hypotheses involving overt behavior (because of the broader framework in which the Os were indoctrinated before the experiment), and the hypothesis is tenable within that context. However, if "personality conflict" is taken to imply subjective reactions, the hypothesis is not supported. High-F persons are probably less sensitive to personality clashes than are low-F, and probably are not as likely to perceive conflicts between themselves and others as are low-F persons.

The Os' ratings of conflict between leaders and group members support the hypothesis that there is less such conflict in homogeneous groups (Table 5). Again, this difference was not found with Ss' ratings of their own groups.

Finally, ratings by Ss and Os yielded conflicting and inconclusive results with regard to hypothesis 5*f*, that homogeneous groups are more productive and have higher morale. With regard to productivity and motivation, nothing very tenable can be concluded. From the observers' point of view, however, it appears that the homogeneous groups had higher morale, higher communication effectiveness, less conflict, and better cooperation than the heterogeneous groups (Table 5).

DISCUSSION

The results, then, support most of the hypotheses tested, indicating significant differences for each of the five major comparisons. Two of these—the comparison of F-plus and F-minus leaders and the comparison of F-plus and F-minus followers—are rather conventional types of comparisons. They simply indicate that the behavior of individuals can be predicted, to some extent, from measures of those individuals' attitudes or personality characteristics.

The other three major comparisons, however, are quite different. Each of these suggests that to some extent the behavior of individuals in groups is dependent on the personalities of other group members. We

found, for example, that there were differences between leaders with F-plus followers and leaders with F-minus followers. This finding indicates that the behavior of leaders is, to a significant degree, a function of the attitudes or personality characteristics of the followers. Conversely, the behavior of followers is found to be significantly a function of the attitudes or personality characteristics of the leaders. The latter, of course, is almost a *sine qua non* for the concept of leadership, but the former has received less attention theoretically and almost none empirically.

Finally, the significant interactions indicate that differences between F-plus and F-minus leaders are contingent on whether the followers are high or low F, and vice versa. This result supports the so-called interaction theory of leadership (8) even more forcefully than the observations referred to in the preceding paragraph. Separating the effects of followers on leader behavior from the interaction effects directs attention to the mutually dependent, adaptive nature of small-group behavior, and appears to facilitate our ability to predict.

The cautious reader will have noted the authors' tendency to reject the null hypothesis occasionally on tenuous bases. There are two relatively noticeable kinds of inconsistencies in the results—those between Session II and Session III, and those between Os' and Ss' ratings. These have been resolved in the discussion by accepting significant results in one session if the other session was not highly contradictory, and to accept Os' ratings as more descriptive of the overt behavior and Ss' as more descriptive of the subjective feelings involved.

Previous evidence (5) has indicated that appointing an individual to a position of leadership in laboratory groups results in his behaving differently from other group members. Another study (10) has shown that group characteristics can be predicted from behavioral and personality measures of individual group members. The present study indicates that conclusions drawn from these previous studies must be tempered by considerations of the interaction among group members, and particularly that between leaders and other group members. All of these findings, of course, are probably contingent on such specifics as the kind of task con-

sidered, the population studied, and the nature of the leadership appointment.

SUMMARY

Four-man groups were required to perform on a task in which Ss discussed specific human relations problems presented by film, then composed and recorded dialogue for similar problems. One of the four men in each group was appointed as leader. The California F scale was administered to all Ss, and group composition was varied systematically to yield four conditions: (a) high-F leaders with high-F followers; (b) high-F leaders with low-F followers; (c) low-F leaders with high-F followers; and (d) low-F leaders with low-F followers. Observations of behavior were recorded by: (a) ratings of individuals on 16 behavioral characteristics, (b) an interaction recording technique using a classification of behavior acts into 43 categories, and (c) responses by Os and Ss to a Post-Meeting Reaction Sheet involving questions about the group.

Five major comparisons were made: (a) F-plus leaders with F-minus leaders; (b) F-plus followers with F-minus followers; (c) leaders of F-plus followers with leaders of F-minus followers; (d) followers with F-plus or F-minus leaders; and (e) groups in which leader and followers were homogeneous as compared to groups in which they were heterogeneous with regard to F-scale scores. For each comparison, significant differences were found consistent with a number of hypotheses drawn from the theory of the "authoritarian personality."

Results indicate that F-plus appointed leaders differ behaviorally from F-minus leaders, and that F-plus followers differ from F-minus followers. These two comparisons are conventional tests of behavioral differences related to personality characteristics. In addition, interaction analyses indicated that the behavior of leaders was a function of whether their followers were F-plus or F-minus, and the behavior of followers depended

on the F-scale scores of their leaders. Several significant individual and group differences were found between heterogeneously and homogeneously composed groups. The implications of these results for an interaction theory of leadership, and for the prediction of behavior from personality measures of individuals, are discussed.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J. & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BLUM, R. *The study of groups*. Washington, D. C.: Human Resources Research Office, Staff Memorandum, 1 August 1953.
3. CARTER, L. F. Leadership and small-group behavior. In M. Sherif & M. W. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953.
4. CARTER, L. F., HAYTHORN, W., MEIROWITZ, BEATRICE, & LANZETTA, J. Note on a new technique of interaction recording. *J. abnorm. soc. Psychol.*, 1951, **46**, 258-260.
5. CARTER, L. F., HAYTHORN, W., SHRIVER, BEATRICE, & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589-595.
6. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. *Handbook for the sixteen personality factor questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1949.
7. DORRIS, R. J., LEVINSON, D. J., & HANFMANN, EUGENIA. Authoritarian personality studied by a new variation of the sentence completion technique. *J. abnorm. soc. Psychol.*, 1954, **49**, 99-108.
8. GIBB, C. A. The research background of an interactional theory of leadership. *Aust. J. Psychol.*, 1950, **2**, 19-42.
9. HAYTHORN, W. The influence of the individual group members on the behavior of co-workers and on the characteristics of groups. Unpublished doctor's dissertation, Univer. of Rochester, 1952.
10. HAYTHORN, W. The influence of individual members on the characteristics of small groups. *J. abnorm. soc. Psychol.*, 1953, **48**, 276-284.
11. HAYTHORN, W., COUCH, A., HAEFNER, D., LANGHAM, P., & CARTER, L. F. The behavior of authoritarian and equalitarian personalities in small groups. *Hum. Relat.*, 1956, **9**, 57-74.
12. STODDILL, R. M. Personal factors associated with leadership: survey of the literature. *J. Psychol.*, 1948, **25**, 35-71.

Received November 15, 1955.

PARADOXICAL EFFECTS IN THE REACTION TIME OF SCHIZOPHRENICS¹

P. H. VENABLES AND J. TIZARD

Medical Research Council, Unit for Research in Occupational Adaptation, Institute of Psychiatry, London, England

IT HAS long been established that for normal, practiced subjects, reaction time to a visual stimulus decreases as the intensity of the stimulus increases (1, 2, 5). Similar findings have been reported for auditory stimuli (3, 9). Hull (6, p. 68), summarizing a number of studies of different kinds, claims that there is a "definite tendency, other things approximately equalised, for trained responses to have a greater reaction potential as the stimulus increases in intensity, whether tested by reaction latency, amplitude, or intensity. Accordingly the principle may be considered as qualitatively well established."

Although among psychologists there has been much discussion about the specific quantitative laws which govern the functional relationship between intensity of stimulus and speed of response, the existence of such a relationship is not disputed. Wundt (12), it is true, is reported by Woodworth (11, p. 118) as saying that RT increases again at high intensities "because one can never be fully ready for a very intense stimulus"; but Woodworth adds that this statement has never been adequately tested. For the intensities usually employed in RT studies, no such increase has been reported.

Pavlov (8) showed, however, that dogs with "weak, inhibitory nervous systems," in whom pathological disturbances of the cortex are induced by means of functional interference, tend to respond to conditioned stimuli in a manner qualitatively different from that of healthy animals. Among the pathological phenomena described by Pavlov are "paradoxical" effects (in which weak stimuli call forth a larger conditioned response than do strong stimuli), and "ultraparadoxical" effects (in which excitatory stimuli are converted into inhibitory ones and inhibitory stimuli into excitatory ones). Stimuli of sufficient strength to bring about paradoxical or ultraparadoxical effects are said to be of "ultramarginal"

strength. Pavlov claimed that schizophrenic withdrawal and negativism are examples of paradoxical and ultraparadoxical phenomena.

Such evidence as there is to support Pavlov's claims is either clinical or based on analogy with experimental neurosis in animals. No experimental studies with human Ss appear to have been carried out. If, however, paradoxical phenomena occur commonly in schizophrenia, it would seem to follow that they should be experimentally demonstrable in a reaction-time study using stimuli of different intensities. Strong stimuli, which ordinarily result in brisk responses by healthy people, should be "ultramarginal" for schizophrenic, and RTs should be slower with such stimuli than with weaker ones.

In a pilot study (10) in which four stimulus lights of different intensities were employed, 21 out of 24 schizophrenic Ss showed such "paradoxical" increases in RT when reacting to lights of higher intensity in contrast to a group of normal Ss whose RTs decreased when they responded to light stimuli of higher intensity. In the present study the factors affecting this paradoxical phenomenon are further investigated under more strictly controlled conditions, with additional Ss and a larger number of stimulus intensities to cover the higher values more fully than had been done before.

METHOD

Procedure. The Ss were tested individually under standard conditions. Eight intensities of stimulus light were used in the experiment, 20 consecutive responses being made to each light.

The stimulus consisted of a ground glass screen 0.75 inches in diameter, illuminated from the rear by a 150 watt bulb fed from a stabilized voltage source. The light source was set in a 2 x 2 feet matt black panel placed at an angle to the vertical so as to be at right angles to the S's line of sight when seated at a distance of 2 feet from the panel. Variations in stimulus intensity were made by inserting a series of neutral density filters behind the ground glass screen. The intensities of the stimuli were measured by means of a photometer; the values obtained are given in Table 1. The background illumination reaching S's eye was 1.7 foot candles.

¹ This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, Dr. H. C. Beccle.

TABLE 1
VISUAL STIMULUS INTENSITIES IN FOOT CANDLES
AND LOG FOOT CANDLES

Intensity	A	B	C	D	E	F	G	H
Foot candles	16	60	135	275	450	600	800	1500
Log foot candles	1.20	1.78	2.13	2.44	2.65	2.78	2.90	3.18

The Ss were tested in two sessions, lights of four of the eight intensities being presented in the first session, and the remaining four in a second session which followed within 24 hours. Order of presentation was varied in a balanced latin-square design such that Ss responded to lights of intensity A, C, E, and G in one session, and to B, D, F, and H in the other. Although 20 RTs were measured for each stimulus intensity, only responses 6-15 were used in the analysis to avoid warm-up and end-spurt effects.

A forewarning buzzer signal preceded by 3.0 seconds the onset of the stimulus light. The S reacted to the stimulus light by lifting his finger. The light then went out and was prevented by a relay mechanism from immediately reappearing when S's finger was replaced on the key to await the next presentation of the stimulus light. If S failed to respond, the light remained on for 3.0 seconds. The cycle of buzzer and light was repeated once every 8 seconds. Between each group of 20 responses, a 3-minute rest period was allowed. The timing of the presentation was controlled electronically to an accuracy of 5 per cent. Measurement of each response was made by means of a cold cathode counter tube timer to the nearest millisecond.

Subjects. All patient Ss were male nonparanoid schizophrenics between the ages of 21 and 52 years. One group consisted of 24 men (mean age, 38.8, *SD*, 8.7 years) who had been in hospital for longer than two years and who could be considered as displaying marked schizophrenic signs. Six out of the 24 had acted as Ss in the pilot study carried out several months previously (10). Eleven other of these *long-stay* patients selected to participate did not cooperate. A second group consisted of 16 convalescent patients (mean age, 28.5, *SD*, 3.6 years) who had been in hospital less than one year, and recently received treatment. This *short-stay* sample was in general much less disturbed than the first group. No S had undergone a leucotomy or was receiving physical treatment or drugs at the time of the experiment.

As controls, four male nurses were also tested. As the pattern of their responses was similar to that already described as typical of normal Ss it was not considered necessary to test a larger control group.

RESULTS

A preliminary analysis showed that the patterns of results obtained from schizophrenic Ss on the first occasion of testing were dissimilar to those obtained on the second. The results from each occasion were therefore analyzed separately. Table 2 shows the mean

RTs for the 12 long-stay patients (Group A), who on the first occasion of testing responded to lights of intensity A, C, E and G, and similar data for the 12 long-stay patients (Group B), who responded to lights of intensity B, D, F and H.

It is seen from these data that, in general, the pattern in both cases is similar, being one of a decrease in RT from the weakest to the next intensity of stimulus (C or D), followed by a subsequent increase in RT with each further increase in stimulus intensity. The decrease in RT from intensity E to intensity G (Group A) is due to the excessively long reaction to intensity E given by two of the 12 Ss in this group. When the data from these two Ss are omitted, the means for this group show a lengthening in RT with each increase in stimulus intensity after the second.

The rise in RT with increase in intensity from C to G is shown by 11 out of 12 Ss, and the rise in RT with increase in intensity from D to H is also shown by 11 out of 12 Ss. Using the sign test (4) as a measure of significance in view of the skewed distribution of the data, the rise in RT with increase in stimulus intensity in each of the groups of 12 Ss is significant at the .01 level. The decrease in RT with increase in intensity from A to C and from B to D is not significant.

Since a similar pattern was found in both groups of 12 Ss, it was considered legitimate to combine the data from each group for the first occasion of testing in order to show how responses varied over the whole range of eight stimulus intensities. As the mean RTs for Group A were higher than those for Group B, those of Group A were proportionately reduced, so that the mean RTs of each group

TABLE 2
MEAN RTs OF LONG-STAY AND SHORT-STAY SCHIZOPHRENIC AND NORMAL Ss ON THE FIRST OCCASION OF TESTING

Subjects	Stimulus Intensity							
	A	B	C	D	E	F	G	H
Long-stay schizophrenics								
Group A	.72		.57		.94		.86	
Group B		.66		.57		.77		.98
Combined group	.69	.66	.55	.57	.91	.77	.83	.98
Short-stay schizophrenics								
Combined group	.34	.35	.31	.32	.30	.34	.32	.31
Normals								
Combined group	.24	.22	.24	.21	.22	.19	.21	.21

were similar. This was done by multiplying the mean RTs of Group A by .965, this figure being derived by division of the summated mean score of Group B by that of Group A. The mean RTs thus obtained for the whole range of intensities presented on the first occasion of testing are shown in Table 2.

The data show that there is a fall in RT from the lowest intensity to the intensities C and D, after which RT rises to a higher figure with increase in intensity.

With the group of short-stay, less severely ill patients, the results obtained for reaction to stimuli of all intensities on the first occasion of testing are shown in Table 2. In view of the closeness of the mean scores of the two subgroups, the one reacting to lights of intensity A, C, E, and G, and the other to B, D, F, and H, no proportional adjustment was necessary. For this group, there were no significant trends in RT with different stimulus intensities.

The mean response times for the four normal Ss for the first occasion of testing are also given in Table 2. Response times for the second occasion of testing were closely similar. They are much shorter than those obtained from the schizophrenics, and like those of the short-stay schizophrenics, they do not show a paradoxical rise with increasing stimulus intensity. The normal group shows only a slight fall in RT as stimulus intensity increases. A more marked fall was not seen because at the intensities employed in the present study, the slope of the curve of performance was approaching the asymptote.

On the second occasion of testing, when the remaining four intensities in the second half of the latin-square design were used, there was no statistically acceptable evidence with either the long- or the short-stay schizophrenic Ss of a change in RT with the increase in stimulus intensity for C and D and beyond. Mean response times over all intensities on the second occasion do not differ significantly from those on the first, although the previously found pattern of response times is absent.

The mean RTs for the long-stay group and for the short-stay group of schizophrenics on the second occasion are given in Table 3.

The results for the six long-stay schizophrenics who had taken part in the pilot study some months before did not differ noticeably from those of the other 18 long-stay patients.

TABLE 3
MEAN RTs OF LONG-STAY AND SHORT-STAY SCHIZOPHRENIC Ss ON THE SECOND OCCASION OF TESTING

Subjects	Stimulus Intensity							
	A	B	C	D	E	F	G	H
Long-stay schizophrenics	.76	.74	.65	.84	.81	.75	.77	.76
Short-stay schizophrenics	.39	.37	.37	.31	.35	.42	.35	.32

In the present experiment five out of the six Ss showed a paradoxical effect on the first but not on the second occasion of testing.

There was no evidence of order effects either within or between testing sessions in either group of patients. Independence of the results in relation to age is shown by a non-significant correlation between age and increase in RT between intensities C and H.

To examine in more detail the nature of the paradoxical phenomenon found with long-stay schizophrenics on the first occasion of testing, further analysis of the data was undertaken. King (7) pointed out that slow RTs are characteristically found among more seriously ill schizophrenics. Though no ratings of degree of illness were made in the present study, the data were examined to see if the paradoxical effect would be shown more markedly by those schizophrenics with the slowest RTs. The total group of long-stay schizophrenics was divided on the basis of RT to the weakest intensity into slow, medium and fast subgroups containing equal numbers of Ss. Linear regression constants, of the form $y = ax + b$, to cover the range from intensity C to intensity H were calculated for each of these subgroups, as well as for the short-stay schizophrenics and the normal controls. These values are given in Table 4.

Table 4 shows that for the slow long-stay schizophrenic Ss, the slope of the gradient of response is steep and positive ($a = +0.872$), but it decreases regularly through a low positive value for the short-stay schizophrenics to a negative value for the normals as the level of RT at intensity C decreases. The negative regression slope shown by the normal Ss in this experiment is in accord with the findings of other workers (1,2,5). The close relationship of regression slope to level of performance at the weakest stimulus intensity is also shown by a correlation of .74 ($p < .001$).

TABLE 4
REGRESSION EQUATIONS SHOWING TRENDS OF
PERFORMANCE OF SCHIZOPHRENIC AND
NORMAL Ss

Long-stay schizo- phrenics	
Slow	$RT = +.872 \log i + .85$
Medium	$RT = +.223 \log i + .46$
Fast	$RT = +.141 \log i + .32$
Short-stay schizo- phrenics	$RT = +.003 \log i + .31$
Normals (male nurses)	$RT = -.029 \log i + .23$

between these two variables based on all the 44 Ss in the study.

A qualitative comment can be made. Some patients were tested who were too ill to cooperate in one or both testing sessions. Their results therefore could not be included in those given above. It was observed however that some of these Ss found it particularly difficult to respond to the brighter lights. Up to the cessation of the light stimulus after the elapse of the 3-second "on" period, they appeared to be pressing down on the key; they were only able to release it when the light went off. Unfortunately, no pressure recordings were made during the experiment, so there is no quantitative confirmation of this impression.

DISCUSSION

The results show with a reasonable degree of clarity that with chronic schizophrenic Ss, the predicted paradoxical effect of an increase in RT with an increase in stimulus intensity beyond an optimum point occurs on initial testing. A problem arising from these data, however, is that the paradoxical effect does not appear on the second occasion of testing. A possible explanation is that on the first occasion, in an unfamiliar situation, Ss were apprehensive and in a state of considerable emotional excitation. The excitation resulting from the stimulus lights, superimposed on this general excitation, became "ultramarginal" in strength and resulted in the paradoxical effect observed. On the second occasion of testing, Ss were more familiar with the situation, and the total excitation did not reach ultramarginal strength. This dependence of the paradoxical effect on the S's level of excitation is in line with Pavlov's views on the

phenomenon (8, pp. 54-56), and it is predictable from his theory that after extensive testing when the general level of excitation of the S is very low because of boredom or fatigue, the paradoxical effect reappears. If subsequent experimentation should prove this to be the case, the second occasion of testing could probably be regarded as an intermediate state between two "phasic" states.

That the paradoxical effect was also shown in five out of six patients who had previously taken part in a pilot study might be considered to weaken the explanation suggested above. But even for these Ss, the situation was probably still a relatively unfamiliar one. The apparatus used differed somewhat from that which had been used in the pilot study; *E* was hardly known to the Ss; and a considerable time had elapsed since their previous testing.

That the phenomenon is fairly well established is shown by its appearance in a total of 56 out of 64 chronic schizophrenics tested in the pilot study (10) already mentioned, and the present investigation.

SUMMARY

A reaction-time experiment was carried out, using lights of different intensities, to test Pavlov's suggestion that paradoxical effects are commonly found in schizophrenics. It was predicted that whereas normal persons tend to respond more quickly to more intense stimuli, schizophrenics respond more slowly to such "ultramarginal" stimuli than to lights of weaker intensity. In 22 out of 24 chronic, nonparanoid schizophrenics, RTs to the brighter stimuli were longer than those to weaker stimuli in a first testing session but not in a second. Convalescent, short-stay schizophrenics and normal controls showed no paradoxical effects. A tentative hypothesis to explain why the chronic patients did not show paradoxical effects on retest is proposed.

REFERENCES

1. BERGER, G. O. Über den Einfluss der Reizstärke auf die Dauer einfacher psychischer Vorgänge mit besonderer Rücksicht auf Lichtreize. *Phil. Stud.*, 1886, 3, 38-93.
2. CATTELL, J. M. The influence of the intensity of the stimulus on the length of the reaction time. *Brain*, 1886, 8, 512-515.
3. CHOCHOLLE, R. Variation des temps de reaction auditifs en fonction de l'intensité a diverses frequences. *Année psychol.*, 1945, 41-42, 65-124.

4. DIXON, W. J., & MOON, A. M. The statistical sign test. *J. Amer. statist. Ass.* 1946, 41, 557-566.
5. FROEBERG, S. The relation between magnitude of the stimulus and the time of the reaction. *Arch. Psychol.*, 1907, No. 8.
6. HULL, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, 56, 67-76.
7. KING, H. E. *Psychomotor aspects of mental disease*. Cambridge: Harvard Univer. Press, 1954.
8. PAVLOV, I. P. *Conditioned reflexes and psychiatry* W. H. Gantt (Trans.). London: Lawrence & Wishart, 1941.
9. PIÉRON, H. De la loi de variation des temps de latence en fonctions des intensités excitatrices par les sensations auditives. *C.R.Soc. Biol.*, 1919, 82, 1116-1118.
10. VENABLES, P. H., & TIZARD, J. The effect of stimulus light intensity on the reaction time of schizophrenics. *Brit. J. Psychol.*, 1956, 47, 144-145.
11. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
12. WUNDT, W. *Grundzuge der physiologischen Psychologie*. (5th Ed.) Leipzig: Englemann, 1911.

Received November 15, 1955.

REACTION TIME AND RESPONSE AMPLITUDE AS A FUNCTION OF ANXIETY AND STIMULUS INTENSITY

ALFRED CASTANEDA

Iowa Child Welfare Research Station, State University of Iowa

ACCORDING to Hull's (2) formulation concerning the joint influence of different drives, total drive strength is assumed to be composed of both irrelevant and relevant drives. Two rather specific consequences are theoretically expected as the result of variations in the magnitude of either type of drive. An increase in one, other factors constant, should result in greater total drive strength. However, increments in total drive strength attributable to an increase in the strength of one are not simply additive but depend on the existing strength of the other. For example, while total drive strength increases with increased irrelevant drive, the increment in total drive strength due to an increase in relevant drive is expected to be relatively greater, the lower the existing strength of the irrelevant drive. The net effect of increasing relevant drive strength under two levels of irrelevant drive, then, is to reduce the difference in total drive strength between the two levels of irrelevant drive. If the situation is relatively simple, then differences in the magnitude of total drive strength should be reflected in the amplitude and latency of the response.

The results of an experiment relevant to these theoretical implications were recently reported by Wenar (8). On the assumptions that differences in intensity of the stimulus in a simple reaction-time task should produce differences in relevant drive, and that anxiety, as inferred from scores on the Taylor Scale of Manifest Anxiety (7), contributes to irrelevant drive strength, the following hypotheses were formulated and tested: (a) the speed of reaction of anxious Ss should be higher than that of non-anxious Ss, and (b) the increase in speed of reaction attributable to increased stimulus intensity should be relatively greater for the nonanxious Ss, thus reducing the difference between the anxious and nonanxious Ss. Results confirming only the first of these two expectations were obtained, i.e., the speed of reaction of the anxious Ss was found to be higher in comparison to those of the nonanxious

Ss, but the difference between them tended to remain relatively constant with increased stimulus intensity.

The present study presents further data relevant to these theoretical expectations. Both the amplitude and the speed of reaction in a simple reaction-time task are used as bases for comparing anxious and nonanxious Ss.

METHOD

Subjects. The Ss were 44 high school boys and girls selected on the basis of scores made on the Taylor scale from a total sample of 152. The distribution of scores for these 152 Ss has been reported by Lyle and Smock (3), who found it to be comparable to that reported by Taylor with over 1,900 college students. The anxious group consisted of 22 Ss, 11 boys and 11 girls, whose scores fell within the range 19-34. The nonanxious group was composed of 22 Ss, 11 boys and 11 girls, whose scores fell within the range 2-12.

Apparatus. The apparatus consisted of a rectangularly shaped plywood box approximately $18 \times 12 \times 6$ in. The response element, the handle and stirrup of a Smedley-type hand dynamometer, projected from the lower right-hand corner of the box and was the only part of the dynamometer visible to S. The CS, a buzzer, was provided by means of a direct current relay energized with alternating current. Variations in the loudness of the buzzer were made possible by means of a variable transformer. Fifteen volts were used for the weak intensity and 45 volts, which produced a fairly raucous sound, for the strong intensity. When the circuit was energized, the buzzer and a Standard Electric clock were simultaneously activated. The indicator on the face of the dynamometer, controlled by the spring-action of the dynamometer, exerted sufficient pressure at the zero position on the leaf of a leaf-type microswitch to keep the circuit closed. Exertion of from four to five kilograms of pressure by S resulted in sufficient deflection of an indicator to open the microswitch and hence break the circuit to the buzzer and the timer. Another indicator remained at the point of maximum pressure exerted by S and had to be returned manually to the zero position by E. Measures of the amplitude of S's reaction as well as of its speed were obtained by this arrangement.

Procedure. The Ss were told that the purpose of the experiment was to measure the speed of their reactions and that they were to squeeze the dynamometer as soon as they heard the buzzer. No information was given regarding the fact that the amplitude of their reactions was to be measured. An interval of seven seconds was used between E's saying "Ready" and the presentation of the buzzer. The E reported S's reac-

tion time to the nearest $\frac{1}{100}$ second after every presentation. The *S* was required to keep his hand on the dynamometer at all times.

The stimulus was presented a total of ten times, five at the weak intensity and five at the loud intensity. The two intensities were randomly distributed over the ten presentations with the one restriction that no one intensity appear more than twice in succession.

RESULTS

Table 1 presents the results based on the amplitude scores by intensity, anxiety level, and sex separately. Table 2 presents a summary of an analysis of variance based on the data presented in Table 1. With the exception of the interaction between sex and anxiety, all other effects are significant.¹ Inspection of Table 1 reveals that the amplitude scores of the anxious *Ss* were higher than those of the nonanxious, a finding consistent with the expectation of greater response strength due to greater irrelevant drive strength. The significant triple interaction, however, when viewed in conjunction with the data presented in Table 1, indicates that the expected greater increase in response strength for the nonanxious group with increased intensity held only for the boys, whereas the performance of the girls was contrary to expectation, i.e., the anxious girls showed the greater increase. Note should be taken of the slight but nonsignificant decrease in amplitude for the anxious boys with increased intensity.

Although the distribution of the latency measures did not appear to be unduly skewed, a reciprocal transformation was performed on all reaction-time scores in order to make them comparable to Wenar's. Hence, the response measure may be referred to as speed of reaction rather than reaction time.

Table 3 presents the data based on the speed-of-reaction scores by intensity, anxiety levels, and sex. On the basis of these data, it is apparent that the speed-of-reaction scores tended to be higher for the nonanxious group at the weak intensity. The increase in speed of reaction associated with increased intensity appeared to be greater in the case of the anxious group. The relatively greater increase in

¹ While differences meet the appropriate tests of statistical significance, it should be noted that there is considerable overlap in the distributions and that the results are not of a magnitude to yield a high degree of accuracy of prediction with respect to individual *Ss*.

TABLE 1
RESPONSE AMPLITUDES IN KILOGRAMS TO THE WEAK
AND LOUD BUZZER FOR THE ANXIOUS AND
NONANXIOUS BOYS AND GIRLS

	Boys				Girls			
	Weak		Loud		Weak		Loud	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Anxious	28.18	5.56	27.27	4.77	17.68	4.52	20.18	3.14
Nonanxious	23.41	8.32	26.50	6.61	15.68	6.02	17.18	5.43

TABLE 2
SUMMARY OF ANALYSIS OF VARIANCE BASED ON
AMPLITUDE SCORES

Source	df	MS	F	p
Anxiety	1	153.02	10.25	<.005
Sex	1	1649.66	110.49	<.001
Sex × Anxiety	1	.31	—	
Error (b)	40	14.93		
Intensity	1	52.65	28.77	<.001
Intensity × Anxiety	1	12.27	6.71	<.01
Intensity × Sex	1	21.22	11.59	<.001
Intensity × Anxiety × Sex	1	17.70	9.67	<.005
Error (w)	172	1.83		

TABLE 3
MEAN OF SPEED-OF-REACTION SCORES AT THE WEAK
AND LOUD BUZZER FOR ANXIOUS AND
NONANXIOUS BOYS AND GIRLS

	Boys				Girls			
	Weak		Loud		Weak		Loud	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Anxious	5.98	1.20	10.29	8.42	4.94	1.60	5.35	1.29
Nonanxious	6.68	2.38	7.54	3.18	4.97	.69	5.08	.62

speed of reaction for the anxious group, despite the tendency for their scores to be lower in comparison to the nonanxious group at the weak intensity, was great enough to result in higher speed-of-reaction scores for these anxious *Ss* at the loud intensity.

Table 4 presents a summary of the analysis of variance performed on the data presented in Table 3. The results of this analysis indicates that with the exception of the main effects of anxiety and the interaction between sex and anxiety, all other effects were statistically significant. Tests of the simple effects of

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE BASED ON
SPEED-OF-REACTION SCORES

Source	df	MS	F	p
Anxiety	1	7.25	2.36	> .10
Sex	1	141.48	46.08	< .001
Anxiety \times Sex	1	4.53	1.48	> .20
Error (b)	40	3.07		
Intensity	1	44.62	15.38	< .001
Intensity \times Anxiety	1	19.29	6.65	< .025
Intensity \times Sex	1	29.87	10.30	< .005
Intensity \times Anxiety \times sex	1	13.69	4.72	< .05
Error (w)	172	2.90		

anxiety, however, did not prove to be significant.

DISCUSSION

Although the results based on the amplitude measure alone tend to be consistent with the assumption of anxiety as an irrelevant drive, they do not lend themselves to an unequivocal interpretation concerning the manner in which stimulus intensity and anxiety combine to determine total drive strength. While the amplitude scores of the anxious group did tend to be higher than those of the nonanxious group at both intensities, the expected reduction in this difference at the loud intensity occurred only in the case of the boys, whereas increased intensity produced an increase in this difference among the girls. On the other hand, the expectation of greater response strength attributable to anxiety, at both intensities, was not reflected in the speed measure and, contrary to expectations, the increase in response strength attributable to increased intensity was relatively greater for the anxious group. If one assumes that the reaction-time task involved only the single motor reaction, then the results based on the speed measure alone are not consistent with the major premises of the present study. It is possible, however, that the task did involve more than the single motor reaction. One may assume that the overt motor reaction was not directly elicited by the CS but was mediated by some implicit verbal response made directly to the CS. In the present experiment, for example, since *S* was instructed to "squeeze" at the onset of the buzzer, the word

"squeeze" may have served to mediate the overt motor reaction of squeezing. If such was the case, it is possible that any other implicit verbal tendencies, e.g., those aroused by the drive stimuli associated with anxiety, could have competed with this mediating response, thereby delaying its occurrence. Assuming the occurrence of the motor reaction to be dependent on the prior occurrence of the mediating response, the effects of other competing tendencies would be manifested in a slowing of the motor reaction. On the other hand, since total effective drive was still operative at this time, the amplitude of the mediated reaction should reflect the magnitude of drive. The present results did show anxiety to be consistently associated with greater amplitude.

In this connection, it should be mentioned that Spence (5) and Postman and Kaplan (4) have stressed the point that measures of speed can provide a sensitive index of the presence of competing tendencies even when other measures fail to do so. If, in addition, it is assumed that the effect of increased stimulus intensity was to increase the strength of the mediating response *relative* to other competing tendencies which may have been present, it would be expected that the weaker the intensity, the greater the tendency for the presence of other competing tendencies to be reflected in the speed measure. By the same token, however, the tendency for differences in the magnitude of drive to be reflected in the speed measure should be greater with increased intensity due to the relative increase in the strength of the mediating response over other competing tendencies. The results reported here with respect to speed are in accord with these speculations. A somewhat similar proposal has been advanced by Spence, Farber, and Taylor (5), who suggest that the difference between anxious and nonanxious *Ss* in classical conditioning situations may be a function of the degree of noxiousness (intensity) in the experimental situation.

Although it might be argued that the present results are consistent with the speculations of an implicit mediating verbal habit and the presence of competing habits associated with the drive stimuli of anxiety, there are no attributes of the present experiment nor of the results which offer any substantially *direct*

evidence in their support. Also, they appear to imply that habits associated with the drive stimuli of anxiety are in some sense incompatible with the mediating *verbal* habit. Such reasoning suggests that some of these interfering habits are also verbal in character. This suggestion accords with the observation that anxious persons are frequently distractible and unable to concentrate, particularly on tasks which require manipulation or retention of symbolic (verbal) material. In fact, the digit-span task is frequently considered as sensitive to the influence of anxiety. Such notions imply rather strong or perseverative implicit verbal habits associated with anxiety, and it is interesting to note that the Taylor scale contains a number of items concerning inability to concentrate, preoccupation with failure, worry, etc. Manifest anxiety symptoms of this type may involve such persistent implicit verbal habits. On the other hand, there are a number of items concerning sweating, heart palpitations, headaches, tension, etc., many of which could be interpreted to indicate that some of the habits associated with anxiety are chiefly somatic or nonverbal. If anxious Ss could be differentiated reliably as to the type of habit associated with anxiety, i.e., whether they are mainly of the verbal or somatic type, their performance might be expected to vary as a function of the type of habit. For example, should the task require a high degree of symbolic manipulation, anxious Ss essentially of the somatic type might perform better than anxious Ss of the predominantly verbal sort. Farber (1) has discussed in detail this stimulus function of anxiety. The present speculations are in line with his considerations in that they suggest a method for the analysis of the habits associated with the stimulus component of anxiety.

SUMMARY

The present study was concerned with reaction time and response amplitude as a

function of CS intensity and anxiety (Taylor scale). Five presentations of two auditory stimuli varying only in intensity were given to each of 22 anxious and 22 nonanxious Ss. Increased intensity resulted in significant increases in amplitude and speed of reaction. Mean amplitude scores of the anxious group were significantly greater than those of the nonanxious group. A significant interaction was found between anxiety and intensity, based on a tendency for the speed of reaction of the anxious group, in comparison to the nonanxious group, to be slower at the weak intensity but faster at the strong intensity. Significant sex differences were found on both measures. The results were discussed in terms of competing implicit response tendencies.

REFERENCES

1. FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311-327.
2. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
3. LYLE, W. H., & SMOCK, C. D. The relationship between manifest and test anxiety. *Proc. Iowa Acad. Sci.*, 1955, **62**, 405-410.
4. POSTMAN, L., & KAPLAN, H. L. Reaction time as a measure of retroactive inhibition. *J. exp. Psychol.*, 1947, **37**, 136-145.
5. SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In *Kentucky symposium, Learning theory, personality theory, and clinical research*. New York: Wiley, 1954. Pp. 1-21.
6. SPENCE, K. W., FARBER, I. E., & TAYLOR, ELAINE. The relation of electric shock and anxiety level to performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404-408.
7. TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
8. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335-340.

Received November 16, 1955.

A GENETIC APPROACH TO THE INTERPRETATION AND EVALUATION OF THE PROCESS-REACTIVE DISTINCTION IN SCHIZOPHRENIA¹

WESLEY C. BECKER

University of Illinois

THE behavior disorders currently labeled "schizophrenia" pose some of the most complex problems psychologists and psychiatrists have yet to unravel. The problem facing the investigator in this area remains an overwhelming one. If controlled research is to be undertaken, some meaningful ways of organizing and understanding the gross individual differences among schizophrenics is necessary.

One compelling suggestion comes from the many prognostic investigations by such people as Hunt and Appel (13), Langfeldt (16), Kant (14), Becker and McFarland (5), Kantor, Wallner, and Winder (15), Stotsky (25), Benjamin (9), and Wittman (26, 27, 28). All of these prognostic studies have led to the conclusion that remitting schizophrenics have a better personality adjustment before and during illness than nonremitting schizophrenics. In addition, these and other studies² have suggested two consistent clusters of signs having differential prognostic significance. Briefly, the signs related to poor prognosis, which for convenience have been termed the "process syndrome," are (a) a "shut-in," withdrawn, inadequate prepsychotic personality, (b) slow, insidious development of psychosis, (c) relative absence of precipitating factors, and (d) presence of dull, rigid, or inappropriate affect. The signs related to good prognosis and termed the "reactive syndrome" are (a) relatively normal prepsychotic personality, (b) acute onset of psychosis, (c) presence of identifiable precipitating factors, and (d) presence of strong emotionality or tension. One implication of these findings is

that a variable of severity of illness or level of adjustment is important to understanding individual differences among schizophrenics.

The consistency of the prognostic findings has led some to postulate the possibility of two kinds of schizophrenia with possible different etiologies. The process syndrome is most frequently assumed to have an "organic" basis, and the reactive syndrome is assumed to have a "psychological" basis (7). However, there are many reasons for rejecting such a conclusion. First, examination of the research data reveals considerable group overlap. Consequently, any attempt to force all schizophrenics into one group or the other would be clinically difficult and arbitrary. Also, if one recognizes with Bellak that schizophrenia is a deficit reaction which may be brought about by any combination of 40 or more etiological factors (8), and if one accepts the fact that twenty years of research have failed to find clear etiological differences between any subgroupings, then the conception of two types of schizophrenia loses usefulness.

An alternative assumption is that the process-reaction syndromes are best thought of as end points on a continuum of levels of personality organization. "Levels of personality organization" is a difficult concept to define succinctly. It is concerned with changes in the content and structure of mental organization as the human organism develops toward psychological maturity. A complete definition would encompass such factors as objectivity in perception (constancy), differentiation of needs, interests, and other aspects of personal motivation, and the degree of emotional control or adaptive functioning under stress. Lewin (17), Baldwin (2), and especially Werner (29) have all attempted to deal with this construct. Common to most attempts to conceptualize levels of personality is the general idea that "the development of biological forms is expressed in an *increasing differentiation* of parts and increasing *subordination*, or *hierarchization*" of the parts with respect to the whole (29). While space

¹ Condensed from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Stanford University. The writer is indebted to Dr. C. L. Winder, Dr. Quinn McNemar, and Dr. Albert Bandura for their guidance throughout this research. He also wishes to thank Dr. Leonard Ullman for his assistance in scoring part of the data, and the director and staff of Agnews State Hospital, Agnew, California, for their helpful cooperation in the execution of this research.

² The reader is referred to Bellak (7) and Kantor *et al.* (15) for reviews of these studies.

does not permit a detailed analysis of the parallel, the general proposition that the process syndrome reflects a very primitive, undifferentiated personality structure and the reactive syndrome a more highly organized one offers a potentially fruitful point of departure.³

To demonstrate the organizing value of this interpretation of the process-reaction concept, and to facilitate measurement of the dimension of personality levels, the following hypothesis was generated and tested: Schizophrenic patients who more nearly approximate the process syndrome show more regressive and immature thinking processes than schizophrenics who more nearly approximate the reactive syndrome. The argument here involves taking a prognostic rating as the criterion measure, placing each schizophrenic *S* along a process-reaction dimension. If this measure is essentially a measure of level of personality organization, and if the hypothesis is sound, then maturity of thought processes should be predictable from the ratings.

METHOD

Newly admitted schizophrenics were interviewed and their clinical records examined in order to gather social history data. This information was then evaluated in terms of the process-reaction dimension. Thinking processes were evaluated from psychological test results. Finally, interrelations between history and test data were explored.⁴

³ A more detailed theoretical exposition can be found in Becker (6). Reference to Werner (29) will also help to make this parallel more explicit for the reader not familiar with Werner's work.

⁴ In considering the method and results which follow, the reader should be aware of a definite methodological difficulty. The author made both the case history and test evaluations. The time-consuming nature of the study did not make it possible to obtain an independent rater for one of the sets of variables. However, definite steps were taken to guard against contamination. Previous studies with the Rorschach have indicated that the scoring system used could be very reliably rated. In addition, an independent reliability study of the Proverbs scoring system was planned within this study. Thus, it seemed unlikely that bias would enter the evaluation of test data, or if it did with regard to the Proverbs, such bias would be detected through a higher correlation with history data by the author than by the independent rater. No such effect was found. It seemed reasonable to assume that if any contamination occurred, it would be more likely in the more complex history ratings. To minimize any such possible effect, those few subscales of the Elgin Prognostic Scale which required direct rating of pa-

Subjects

The *Ss* were 51 diagnosed schizophrenics selected at time of admission to a California state hospital (Agnews) on the basis of their being testable, non-Negro, under 41 years of age, without complicating neurological or other physical disorder, and on the basis of their being either a first admission case or a second admission case where the first admission was under four months duration. Twenty-four *Ss* were male, 27 female. Average age for the men was 29.4 and for the women, 31.1. Twenty per cent had attended college, and over 60 per cent had been graduated from high school. About half were married. All but seven cases were first admissions. Twenty-four were diagnosed as paranoid type, nine as acute undifferentiated type, eight as chronic undifferentiated type, eight as schizo-affective type, and one each as catatonic and hebephrenic.

Evaluation of Thinking Processes

The independent variable to be evaluated is degree of regressive and immature thinking. In the literature one finds many definitions of schizophrenic thinking disorder, but few, if any, of such definitions contain the possibility of scaled quantification. The conceptual structure taken in this paper, however, suggests both theoretical and empirical (i.e., studies with children) methods of scaling severity of thinking disorder by initially equating it with immaturity of thinking.⁵ To

tient behavior were rated before testing. In addition, all 51 cases were interviewed and tested before other case-history evaluations were made. In the meantime, no attempt whatsoever was made to evaluate the test data. For this reason, by the time the author was ready to make the final evaluations of case histories from the clinical records—a time lapse of from one to four months from testing—it was close to impossible to associate a specific test protocol with a specific case history. While one cannot assert that no bias entered into the ratings, the rather high reliabilities of all measures used and the precautions taken make it unlikely.

Ideally a follow-up study using independent evaluations would be desirable. Such a study is being planned under the direction of C. L. Winder as part of a comprehensive five-year study. Because of the length of time before the results of Dr. Winder's study will be available and because the author believes this paper has definite methodological and theoretical contributions to offer which are independent of the results, publication was sought at what may seem in some ways to be a premature date.

⁵ To avoid misunderstanding, elaboration of this point is needed. What is asserted here is that the *formal organization* of mental processes of severe schizophrenics and young children, or of less severe schizophrenics and slightly older children is quite similar. The assumed basis for this similarity is that the schizophrenic has been fixated at an immature level of personality organization and/or has reverted to simpler levels of organization under stress. This approach provides a rational framework for the initial scaling of

demonstrate the method and to test the hypothesis given above, the Rorschach and Benjamin Proverbs were used to evaluate thinking processes. The 1937 Stanford-Binet Vocabulary test was included in order to estimate verbal intelligence.

Rorschach. Administration of the Rorschach followed the procedure proposed by Beck (3, 4). The method of analyzing the Rorschach was derived most directly from Werner's developmental theories (29) and the empirical studies by Friedman (10), Siegel (23), and Hemmendinger (12). In studying the problem of perceptual regression in schizophrenia, Friedman developed a scoring system for the Rorschach which would presumably reflect levels of perceptual development as suggested by Werner's theory. In comparing adult normals, schizophrenics, and children of various ages on his measures, he found strong support for the hypothesis that the structural aspects of schizophrenic perception were in many ways similar to, though not identical with those of children. Werner's conceptual structure and these supporting studies encouraged the notion that a scale could be devised which would differentiate levels of mental organization within the schizophrenic population.

In quantifying his data, Friedman used two gross categories—genetic-early and genetic-late perceptions. Each genetic group was further divided into three progressive levels, giving a total of six levels. The empirical findings of Friedman and Siegel, as well as Werner's theories were employed in making these differentiations. In order to derive an over-all score to reflect average level of mental organization, each response was given a weight from one to six, corresponding to the genetic level of response, and the sum of the weights was divided by the number of responses.

A summary outline of the Rorschach scoring system with examples is given in Table 1.⁶ Rationale for placement of categories is discussed below.

Level one is characterized by diffuse, global, undifferentiated perceptions. *Wa* is a direct indicator of the diffuse, global nature of perception at this level. *W-* indicates the lack of differentiation of inner and outer worlds and the syncretism in mental organization. *DW* is a clear example of a type of diffuse perception Werner has described by the phrase "*pars pro toto*"—any part has the quality of the whole (29). The contaminated

schizophrenic mental productions. Empirical findings, such as those by Friedman (10), can then be used to refine such a scale and to take into account any differences which may exist in the mental productions of children and schizophrenics of a given organizational level.

⁶ The reader is referred to Friedman (10) and Siegel (23) for more complete definitions of scoring categories, with the exception of Unusual Details and Perseveration. The Unusual Detail scores are discussed in the text. Friedman's use of a perseveration score based on Beck's content classes did not prove to be discriminatory between age groups. For this reason the writer adopted Piotrowski's (21) more severe criterion for perseveration. Examples in the table are from Friedman (10), Phillips and Smith (20), and the author's own experience.

and fabulized responses reflect both the concrete and syncretic nature of primitive perception. In these responses there is an absurd fusion of percepts on the basis of spatial identity and spatial contiguity. Perseveration is assumed to reflect the dynamic rigidity of the personality at this level. Empirically, Friedman, Siegel, and Hemmendinger found these perceptual classifications most characteristic of children under five years of age and of severely regressed schizophrenics.

Level two is characterized by attempted differentiation in which the diffuse and syncretic nature of perception is still apparent. Friedman found the *Da*, *DdD*, and *Dv* responses all relatively rare in children, suggesting that where perceptual organization is so advanced as to permit discreteness (*D*), the diffuse and syncretic modes of perception have already begun to wane. In psychopathology, however, one might expect to find a greater frequency of this combination of discreteness with diffusion and syncretism where there is regression from higher levels of perceptual differentiation. A similar explanation is offered by Friedman to account for the differences in relative frequencies of the *Wv* and *Dv* responses. The greater frequency of vague responses in the schizophrenic group may be the result of regressive diffuseness and vagueness interfering with percepts which might have been *Wm* responses. Because *Wv* shows some integrative effort with consideration of the formal aspects of the blot it is placed at level three. *Dv* lacks integrative effort and is therefore placed at level two. *D-* and *Dd-* again are indicative of an unsuccessful attempt at differentiated perception.

Level three is characterized by the achievement of fair differentiation with only rudimentary integrative efforts. Developmentally, the scores at this level (with the exception of *Wv*) are most characteristic of children from ages seven to ten. In some ways *Adx-Hdx* responses are like the *Dm* of the next level. However, they are placed at this lower level because they clearly indicate a failure at integration where inegration is usually easily achieved. *Dd* responses were omitted from Friedman's schema. However, Hemmendinger (12) and Siegel (23) indicate that *Dd*'s are found most frequently between the ages of six and ten. Within this age range, an average of 13 per cent of the responses were of the *Dd* variety. None was found before the age of six, and normal adults gave but 4 per cent *Dd*. Siegel's less severe paranoid group gave 15 per cent *Dd*, whereas his more severe hebephrenics and catatonics gave but 3 per cent. These data are consistent with the interpretation that *Dd* is most characteristic of a level of development where the primary focus is analytical. With Beck's tables as a guide (3, 4), *Dd+* was placed at level three, while the more immature *Dd-* was placed at level two.

Level four is a stage of accurate differentiation with the ability to make simple integrations. The Mediocre responses scored at this level indicate the ability to meet certain constant typical requirements in form necessary for adult perception.

Level five is indicative of clearly integrative activity with the ability to subordinate differentiated parts to the whole.

TABLE 1
DEFINITIONS AND EXAMPLES FOR THE RORSCHACH GENETIC-LEVEL SCORING SYSTEM

Level	Classification	Definition	Examples
1	Amorphous Whole (<i>W_a</i>)	Shape plays no determinable role.	I. "Black paint"
	Minus Whole (<i>W₋</i>)	Content requires specific form not provided by blot.*	II. "Fire and smoke"
	Confabulatory Response (<i>DW</i>)	A single detail is basis for interpretation of the whole.	I. "A fly"
	Contaminated Response (<i>Con R</i>)	Fusing of two interpretations of the same blot area.	IV. "Starfish"
	Fabulized Combination (<i>Fab C</i>)	Absurd combination on basis of spatial contiguity.	VI. "Cat," because of "whiskers"
	Perseveration (<i>Per</i>)	Same content to 3 or more cards with little regard to form requirements.	VI. "Turtle-skin rug"
2	Amorphous Detail (<i>Da</i>)	Analogous to <i>W_a</i> .	X. "Rabbit with worms coming out of eyes"
	Confabulatory Detail (<i>DdD</i>)	Analogous to <i>DW</i> .	I, IV, V. "Spider"
	Minus Detail (<i>D₋</i>)	Analogous to <i>W₋</i> .*	VIII, IX, X. "Internal organs"
	Vague Detail (<i>Dv</i>)	Form element is so unspecific that almost any blot area could encompass content.	II. (<i>D</i> 2) "Fire"
	Minus Unusual Detail (<i>Dd₋</i>)	Analogous to <i>W₋</i> .*	VIII. (<i>D</i> 6) "Flesh"
3	Vague Whole (<i>W_v</i>)	Analogous to <i>Dv</i> .	VI. (<i>D</i> 3) "Cat's head," solely on "whiskers"
	Oligophrenic Detail (<i>Adx-Hdx</i>)	Response to part of an <i>A</i> or <i>H</i> percept usually seen as a completed figure.	II. (<i>D</i> 2) "Kittens"
	Plus Unusual Detail (<i>Dd₊</i>)	Content is a reasonable match to blot area isolated.*	II. (<i>D</i> 2, 3) "Blood stains"
			X. (<i>D</i> 9) "Island"
4	Mediocre Detail (<i>Dm</i>)	Form implied in outline and articulation matches blot area. At level of "populars."	VI. (<i>Dd</i> 25) "Pig's foot"
	Mediocre Whole (<i>W_m</i>)	Analogous to <i>Dm</i> , but applies only to unbroken blots.†	I. "Piece of a puzzle"
			X. "Design," "Map"
5	Plus Detail (<i>D₊</i>)	Two or more <i>D</i> areas are combined into one "good form" percept.*	III. (<i>D</i> 6) "Head of a person"
	Plus Whole (<i>W₊</i>)	All <i>D</i> portions of a broken plot are combined into one "good form" percept.*†	V. (<i>D</i> 4) "Wing"
			X. (<i>D</i> 26) "Funny face"
6	Plus-Plus Whole (<i>W₊₊</i>)	An unbroken blot is perceptually articulated and reintegrated into a "good form" percent.†	III. (<i>D</i> 3) "Bow tie"
	Plus-Plus Detail (<i>D₊₊</i>)	A <i>D</i> area is articulated and reintegrated into a "good form" percept.	X. (<i>D</i> 15) "Little bird"
			I. "Bat," "fox's head"
			VI. "Mud turtle"
			II. (<i>D</i> 1's) "Bears fighting"
			II. "Two fellows at a bar toasting each other"
			IV. "A giant sitting on a stump"
			X. (<i>D</i> 8, left) "Guy riding a horse"

* Beck's tables (3,4) are used as a guide in scoring.

† Unbroken blots are I, IV, V, VI, and IX; broken blots are II, III, VII, VIII, and X.

Level six is characterized by the highest form of differentiation and hierarchic integration which is found only in mature perception.

In scoring, additional responses given during the inquiry were included. If *S* rejected a response on inquiry, it was excluded. If *S* improved a response on

inquiry, he was given appropriate credit. The focus was on how well *S* could do rather than on how poorly. Beck's locations (3, 4) were used for *D* and *Dd* areas. Siegel (23) found the mean percentage of agreement among three judges to be 93.9 per cent, using all of the above scoring categories with the exception of

perseveration and the distinction between *Dd+* and *Dd-*. His findings indicated that the definitions are clear enough to be used by other Rorschach examiners. No further reliability studies were made at this time.

Proverbs test. Benjamin (9) was one of the first to apply proverbs in a systematic way to the study of schizophrenic thinking. As Werner's theories would suggest, the poorly differentiated schizophrenic and the child tend to take a literal or concrete approach to proverb interpretation. In administering the test, S was given an example and told to "interpret the proverb, to try to give a general meaning." Thirteen of the 14 proverbs were then given in the order originally used by Benjamin (9). Proverb 13 (an absurd phrase and not a proverb) was omitted. While Benjamin has never published a systematic scoring system for his test, he has described the types of thinking pathology exhibited through proverb interpretation (9). In developing the scoring system, both Benjamin's analysis and Werner's theories were taken into consideration. The following general scoring classifications were used⁷:

Abstract III: A correct generalized interpretation without detracting elements. (Weight of 6.) Example: "Brooding over past mistakes is futile."

Abstract II: A correct example with reference to human behavior; another proverb meaning the same thing; a response partly generalized, partly restricted to a specific example; a lower level generalization. (Weight of 5.) Example: "What's done is done."

Abstract I: A response tinged with the literal; a response which would be acceptable at Abstract II but for some minor inaccuracy, overstatement, or alternative explanation which is false. (Weight of 4.) Example: "Don't cry over something that's happened; can't be helped" (literal tinge).

Vague Response: An attempt at interpretation which is on the right track, but is left too vague to be adequate, or fails to account for part of the proverb. (Weight of 3.) Example: "It's too late, in other words."

False Interpretation: The interpretation is very inaccurate, yet an attempt was made to interpret. The error is usually due to faulty desymbolization or faulty generalization. (Weight of 2.) Example: "Don't let defeat stop you."

General Literal: The interpretation is literal in effect, though stated in general terms. At first glance these responses do not appear to be literal, but they can only be understood as stemming from a literal interpretation. (Weight of 2.) Example: "What is wasted is wasted; no tears."

Literal: The proverb is interpreted literally. (Weight of 2.) Example: "Don't cry when you spill some milk."

Absurd: The response indicates a failure to interpret and/or is logically absurd in terms of the task at hand. These responses are usually based on concrete associations to some aspects of the stimulus. (Weight of 1.) Example: "The milk is on the floor and the horses will drink the water."

Literal-abstract: A response which gives both a literal and an abstract interpretation. The tendency to be drawn into a literal interpretation is strong, but the subject is able to counteract it. (Weight at best abstract level reached.) Example: "Call the cat to lick it up, or what is happened has happened; let's look ahead."

To determine rater reliability, 25 cases were re-scored by a fourth-year clinical psychology student. This second rater used the general definitions given above, plus examples drawn from the cases not re-scored. The two-rater reliability was found to be .98. Odd-even reliability for 50 cases was found to be .83. It would appear that the scoring system is reliable and that the different parts of the Proverbs test define a similar function.

Because of a high relationship between ability to interpret proverbs and verbal intelligence in normals (18), it was assumed that a more sensitive index of thinking disturbance would be a discrepancy score based on the standard-score difference between a vocabulary estimate of verbal intelligence and the Proverbs score.^{8, 9}

⁸ The adequacy of this assumption is tested later. The usual argument is that the degree of disturbance in mental organization is best measured by the loss relative to the individual's own potential rather than loss relative to an average potential. Babcock (1), Shipley (22), and Simmins (24) have all supported the pragmatic value of this contention. However, this would imply that mental maturity is a relative concept, whereas in scoring the Rorschach, maturity levels are generally defined in terms of absolute standards. This dilemma seems to have arisen because of differential knowledge in the areas of perceptual organization and conceptual organization. In the former case, there are no ready tests and data suggesting how well a person might do under optimal conditions (*i.e.*, without psychosis), so judgments of severity are made on the basis of an average expectation (absolute standards). In the latter case, there are ways of estimating one's potential, and this information can be used to gain a more refined prediction. The basic source of difficulty seems to lie in the equation of immaturity with severity of illness. Perhaps immaturity-maturity might better be taken as the reference scale and severity of illness defined as the difference on this scale between potential and present levels of functioning. This problem needs further thought and study.

⁹ The actual measure of vocabulary level used was the last word right before six consecutive failures. This measure correlates .907 with the usual way of scoring vocabulary, but seems to give a better estimate of potential verbal intelligence for the acutely disturbed and chronically withdrawn, who show greater inconsistency in response to any verbal test.

⁷ The reader is referred to the original work (6) for more detailed scoring instructions and for examples to be used as a guide in scoring. The examples given here are taken from the second proverb, "Don't cry over spilt milk." It should be noted that this scoring system does not have the empirical foundation that Friedman's and Siegel's studies gave to the Rorschach. Genetic studies might help to improve this system.

Evaluation of Process-Reaction from Case Histories

The method of quantifying the social history data so as to reflect the process-reaction dimension was adapted from Wittman (26, 27). Wittman has made repeated studies of the prognostic value of case history data and formalized her findings as the Elgin Prognostic Scale. The scale consists of 20 variables, weighted according to prognostic significance. Included in the subscales are evaluations of prepsychotic personality, nature of onset, and typicality of the psychosis relative to Kraepelin's definition. These are the same factors on which the process-reaction distinction was made, so that Wittman's scale is an adequate summary of what has typically been included in the process-reaction distinction.

Three primary sources of data were used in rating the Elgin scale: (a) direct interview with the patient, (b) anamnesis, and (c) psychiatric history obtained by the examining psychiatrist. Using her scale with similar sources of information, Wittman found a two-rater reliability of .87 for 61 cases (26). In several studies of the prognostic validity of the scale, Wittman (26, 27, 28) found the scale to predict outcome, with or without shock therapy, in 80 to 85 per cent of the cases. These data indicate that this is a moderately reliable instrument for assessing the process-reaction dimension.

Wittman has defined only the end points of each subscale. To add precision to the ratings, definitions were developed for intermediate points.¹⁰ The added definitions adhered closely to Wittman's original intent as to subscale meaning. Though a check of the reliability of the modified scale was not possible, it is doubtful if it would be less than that reported by Wittman.

PREDICTIONS

The following specifications of the hypothesis were made: Those schizophrenics with more process-like case histories (higher scores on the Elgin scale) receive significantly lower mean-genetic-level scores on the Rorschach and significantly lower Proverbs-minus-vocabulary discrepancy scores than those schizophrenics with more reactive-like case histories. The critical acceptance level was set at .01, using one-tailed tests of significance.

RESULTS

The distribution of scores on the Elgin scale were first normalized by use of McCall's *T*-score method (11). Analysis showed a significant difference (.02 level) in the means for the men and women on the Elgin scale. The mean for the men was 52.9 and for the women 46.2. This difference suggests that on

the average the women in this sample were less severe cases. A possible selective factor is the care given a large portion of the less severe males by the Veterans Administration hospitals in the San Francisco area. Because of this significant difference in means on the Elgin scale, the hypotheses were first examined separately by sex group. If the relationships did not differ significantly, the correlations were then averaged to obtain an estimate of the population r independent of the sex difference in means on the Elgin scale.

Main Results

The Rorschach hypothesis was confirmed for both sex groups. The Rorschach mean-genetic-level score and the Elgin Prognostic Scale correlated $-.599$ ($p < .01$) for the men, and $-.679$ ($p < .001$) for the women. Since these correlations did not differ significantly, an averaged r was computed and found to be $-.641$ ($p < .001$).¹¹ These results indicate that there is a meaningful relationship between the process-reaction dimension, as evaluated from case-history data, and disturbances of thought processes as measured by the Rorschach genetic-level scoring system.

The hypothesis concerning the Proverbs test was confirmed only for male schizophrenics. The correlation between the Proverbs discrepancy score and the Elgin scale was $-.682$ ($p < .001$) for men, and .048 (obviously insignificant) for the women. The difference between these two correlations is significant beyond the .01 level. These results indicate that the Proverbs-vocabulary discrepancy score is significantly related to the process-reaction dimension for men, but not for women.

Further Analyses

In making further analyses, certain interrelations of the data were explored. First, it was found that the Rorschach and Proverbs-vocabulary discrepancy scores measured a similar function for the men ($r = .599$; $p < .01$) but not for the women ($r = -.210$). Second, the adequacy of the assumptions about vocabulary corrections was tested, using a multiple correlation technique. No predictive gain was found when vocabulary was used in the re-

¹⁰ See Appendix I in Becker (6).

¹¹ Fisher's z transformation was used to compute the averaged r (19).

gression equation to predict Elgin-scale placement from the Rorschach. This result is consistent with the initial assumption.¹² A similar analysis for the Proverbs test (meaningful only for the men) showed an increase in correlation from $-.603$ for Proverbs alone to $.690$ (corrected for "shrinkage") when both Proverbs and vocabulary were used to predict the Elgin scale.¹³ The multiple R gives about the same result as was reported earlier for discrepancy scores ($-.682$) and supports the value of a vocabulary correction. Finally, the effectiveness of the best linear combination of the Rorschach, Proverbs, and vocabulary scores in predicting the Elgin scale was investigated. For the men, the multiple R was $.768$ after correction for "shrinkage." All three tests contributed substantially to the prediction with vocabulary acting as a suppressor variable. For the women, the multiple R showed no increase over a direct prediction from the Rorschach, since the Proverbs test carries no weight for this group.

DISCUSSION

The results in part support the central hypothesis and demonstrate the value of a levels-of-personality interpretation of the process-reaction construct. A number of questions, however, were raised by the findings. Most difficult to understand is the sex difference in the meaning of the Proverbs test. Elaborate statistical analyses of the data failed to clarify this situation. A possible clue may lie in the fact that predictions were confirmed as long as the focus was on perceptual organization (Rorschach) but failed to be confirmed for women with the shift to conceptual organization (Proverbs). Possibly conceptual functioning in schizophrenic women is more easily disrupted than perceptual functioning by momentary emotional factors independent of severity of illness. However one interprets these results, if supported by future research, they imply a need either to control for the "interfering factor" or to develop different types of instruments for the

evaluation of degree of pathology in male and female schizophrenics.

Two further implications should be mentioned. First, the Rorschach findings may be interpreted as evidence for the validity of the Rorschach in reflecting levels of personality organization. In addition, a theoretical framework (seldom found in the Rorschach literature) is offered for why the Rorschach should be able to perform such a task. Second, the coherence of the findings suggests the feasibility of developing a diagnostic system similar to the stanine system used by the Air Force. On the basis of history and test data, schizophrenics could be classified as Level One's, Level Two's, etc., and the diagnosis would have genuinely prognostic value. In addition, such a classification would provide a sounder basis for distinguishing subgroups for etiological research, an important control variable for research on the effectiveness of therapy, and a frame of reference for evaluating the course of the disturbance over time.

SUMMARY

The Elgin Prognostic Scale was used to evaluate the case records of 24 male and 27 female schizophrenics in terms of a process-reaction continuum on the assumption of its equivalence to a measure of level of personality organization. Werner's developmental theories and certain empirical studies were used to score Rorschach and Benjamin Proverbs responses in terms of levels of differentiation and integration. It was then predicted that more process-like schizophrenics obtain lower Rorschach genetic-level scores and lower Proverbs-minus-vocabulary scores than more reactive-like schizophrenics. Because of a significant difference between men and women on the Elgin scale, the hypotheses were examined separately for the sex groups.

The Rorschach hypothesis was confirmed for both sex groups. The r between the Rorschach genetic-level score and the Elgin Prognostic Scale was $-.599$ ($p < .01$) for the men and $-.679$ ($p < .001$) for the women. The hypothesis for the Proverbs test was confirmed only for male schizophrenics. The r between the Proverbs discrepancy score and the Elgin scale was $-.682$ ($p < .001$) for the men and $.048$ for the women. No adequate explanation for this sex difference was found.

¹² A correction based on some measure of potential level of perceptual organization might add to the predictive efficiency of the Rorschach test.

¹³ All multiple R s are positive in sign. The implied relationship in this case is the same as if the multiple R had a negative sign.

The general conclusion is drawn that the results in part support the central hypothesis. There is evidence for a measurable dimension of regressive and immature thinking which is related to the process-reaction dimension. In addition, the value of a levels-of-personality-organization interpretation of the process-reaction dimension is supported.

REFERENCES

1. BABCOCK, HARRIET. *Dementia praecox, a psychological study*. Lancaster, Pa.: Science Press, 1933.
2. BALDWIN, A. L. *Behavior and development in childhood*. New York: Dryden, 1955.
3. BECK, S. J. *Rorschach's test*. Vol. I. *Basic processes*. New York: Grune & Stratton, 1944.
4. BECK, S. J. *Rorschach's test*. Vol. I. *Basic processes*. (2nd Ed.) New York: Grune & Stratton, 1949.
5. BECKER, W. C., & MCFARLAND, R. L. A lobotomy prognosis scale. *J. consult. Psychol.*, 1955, **19**, 157-162.
6. BECKER, W. C. The relation of severity of thinking disorder to the process-reactive concept of schizophrenia. Unpublished doctor's dissertation, Stanford Univer., 1955.
7. BELLAK, L. *Dementia praecox*. New York: Grune & Stratton, 1948.
8. BELLAK, L. A multiple-factor psychosomatic theory of schizophrenia. *Psychiat. quart.*, 1949, **23**, 738-755.
9. BENJAMIN, J. D. A method for distinguishing and evaluating formal thinking disorders in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia*. Berkeley: Univer. of California Press, 1946. Pp. 66-71.
10. FRIEDMAN, H. Perceptual regression in schizophrenia. An hypothesis suggested by use of the Rorschach test. *J. proj. Tech.*, 1953, **17**, 171-185.
11. GUILFORD, J. P. *Fundamental statistics in psychology and education*. (2nd Ed.) New York: McGraw-Hill, 1950.
12. HEMMENDINGER, L. Perceptual organization and development as reflected in the structure of Rorschach test responses. *J. proj. Tech.*, 1953, **17**, 162-170.
13. HUNT, R. C., & APPEL, K. E. Prognosis in psychoses lying midway between schizophrenia and manic-depressive psychoses. *Amer. J. Psychiat.*, 1936, **93**, 313-339.
14. KANT, O. Differential diagnosis of schizophrenia in the light of concepts of personality stratification. *Amer. J. Psychiat.*, 1940, **97**, 342-357.
15. KANTOR, R. E., WALLNER, J. M., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157-162.
16. LANGFELDT, G. Prognosis in schizophrenia and factors influencing course of disease: catamnestic study, including individual re-examinations in 1936 with some considerations regarding diagnosis, pathogenesis and therapy. *Acta Psychiat. Neurol.*, 1937, Suppl. 13. Pp. 1-228.
17. LEWIN, K. *A dynamic theory of personality*. D. K. Adams & K. E. Zener (Trans.). New York: McGraw-Hill, 1935.
18. MCNEMAR, Q. *The revision of the Stanford-Binet scale*. New York: Houghton Mifflin, 1942.
19. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
20. PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: advanced technique*. New York: Grune & Stratton, 1953.
21. PIOTROWSKI, Z. A. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.*, 1936-1937, **1**, 65-77.
22. SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, **9**, 371-377.
23. SIEGEL, E. L. Genetic parallels of perceptual structuralization in paranoid schizophrenia: an analysis by means of the Rorschach technique. *J. proj. Tech.*, 1953, **17**, 151-161.
24. SIMMINS, C. Studies in experimental psychiatry: IV. Deterioration of 'G' in psychotic patients. *J. ment. Sci.*, 1933, **79**, 704-734.
25. STOTSKY, B. A. A comparison of remitting and nonremitting schizophrenics on psychological tests. *J. abnorm. soc. Psychol.*, 1952, **47**, 489-496.
26. WITTMAN, PHYLLIS. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hosp. Papers*, 1941, **4**, 20-33.
27. WITTMAN, PHYLLIS. Follow-up on Elgin prognosis scale results. *Illinois psychiat. J.*, 1944, **4**, 56-59.
28. WITTMAN, PHYLLIS, & STEINBERG, D. L. Follow-up of objective evaluation. *Elgin State Hosp. Papers*, 1944, **5**, 216-227.
29. WERNER, H. *The comparative psychology of mental development*. (Rev. Ed.) Chicago: Follet, 1948.

Received November 21, 1955.

ASSESSMENT OF ABSTRACT BEHAVIOR IN A NON-WESTERN CULTURE

GUSTAV JAHODA

University College of the Gold Coast

CROSS-CULTURAL studies of test intelligence, and critiques of such studies abound in the literature and have been summarized by Anastasi and Foley (1), but abstract behavior or its alleged absence more often forms the subject of impressionistic judgments. Carothers discusses some of these with special reference to Africans and does not himself dissent from the view that the African "is inapt for sound abstraction and logic" (9, p. 87). Another recent example of sweeping generalization in this field is to be found in a series of papers by Haward and Roland (13, 14, 15), reporting the results of administering the Goodenough Draw-a-Man Test to some inadequately specified samples of Nigerians. Apparently not fully aware of the limitations of this test when used in non-Western cultures (12), these authors write about "the concreteness of the Nigerian's mental approach, a concreteness which is so rigid as to produce schizophrenic signs" (13, p. 88).

Until the work of McConnell (20) no attempt appears to have been made to assess the capacity for abstract behavior of non-Western people by means of the tests specially devised and commonly used in Western countries for this purpose. It was not the aim of the present study to discover the extent to which the unflattering statements about the African's "concrete mentality" would seem to be confirmed or contradicted by such a measure as the Goldstein-Scheerer Cube Test (GSCT); but rather to examine the question whether a test of this kind is any more appropriate for making cross-cultural comparisons than various types of intelligence tests, some of which it closely resembles.

METHOD

Subjects

The work was carried out in three stages, with independent groups of adolescent boys. Except for a small minority from other parts of the West Coast they were Gold Coast Africans. Over two-thirds belonged to the Ga tribe, indigenous to the Accra region; the remainder were mainly the sons of Ewe or Akan mi-

grants. In the absence of universally compulsory registration of births, ages are difficult to ascertain; they ranged from 13 to about 18 years, but the modal ages varied somewhat. All Ss were attending middle schools and had received eight or nine years of elementary education.¹

The first group, whose modal age was about 15 years, took the GSCT. The group consisted of 27 boys from the Accra urban area, representing the remnant of an original sample of 33,² which had been selected in the following way: on the basis of information supplied by the education authorities concerning geographical and social "catchment areas" two schools were chosen as appearing to offer jointly a fairly satisfactory cross section of the middle-school population. Within the schools the third or penultimate form was decided upon after balancing the considerations of adequate command of English vs. duration of availability for investigation. Although some work could be done with the complete forms, for the purpose of intensive study two subsamples of 18 and 15 boys respectively were drawn by means of a table of random numbers.

At the time of the GSCT the investigator had been in contact with these boys for nearly four months. The home of each one had been visited, so that it was possible to decide reliably whether their home backgrounds were literate or illiterate. To ascertain this is a less simple matter than might be imagined, and the mere knowledge of the father's educational level is insufficient; children of a literate father may spend their lives with an illiterate mother, perhaps the second wife in a polygynous family, or with other illiterate relatives.

The second sample consisted of ten complete forms,

¹ It would be desirable to relate the school population from which these samples were drawn to the entire age groups. Unfortunately no adequate statistical data about age distribution in the general population is available. An estimate has, however, been published (28) of the total number of children of middle-school age in the Gold Coast as a whole. The sex ratio being known (30) it is possible to compare this estimate with the figures of boys' middle-school enrollment (29). From this one can roughly assess the proportion of eligible boys who attended middle schools at the time of the investigation as just under a half (it may be of interest that the proportion was only a tenth in the case of girls). In the Accra region the rate of enrollment was probably substantially higher. In the writer's view the original estimate is subject to a very wide margin of error; the inferences based on it are therefore hardly more than informed guesses.

² Two went to secondary schools; two more failed and had to leave; one moved to another town, and one became delinquent and was expelled.

numbering over four hundred, from several urban schools. These were given Raven's Progressive Matrices Test (23) on three successive weeks, and 317 boys completed the whole series. Their modal age was approximately 16 years.³

While the investigator was working in a semirural school he came across the work of McConnell (20). It was possible to replicate his study with one complete form of 32 boys, whose modal age was again about 15 years.

The medium of instruction in middle schools is English, but practically all the boys speak only the vernacular at home. Great care was therefore taken to ensure that test instructions were clearly understood.

Tests Employed

Goldstein-Scheerer Cube Test. The terms "abstract" and "concrete" are sometimes used as if they were descriptive of two contrasting types of mentality. In fact, as with nearly all psychological variables, one is dealing with a continuum. In their monograph (11, p. 48 ff.) Goldstein and Scheerer claim that the GSCT can serve to differentiate between abstract and concrete types of normals, and it thus appeared a suitable tool for exploring the problem. Further advantages are the quantitative scoring system and rough norms devised by Boyd (8), which facilitate comparison.

The question of whether the Cube Test would be valid in other cultures was not explicitly faced by Goldstein and Scheerer. It must be remembered that they are primarily concerned with the diagnosis of pathological impairment, and "normal" in their usage probably has to be understood in this sense. They do state quite clearly, however, that all "normals" are equipped to solve the Cube tests, at any rate with the aids provided.⁴ It would seem to follow that "normal" people in other cultures should also be able to perform successfully.

Some of Goldstein and Scheerer's underlying assumptions will be examined later, but it might be mentioned at this stage that Boyd (8) was able to demonstrate a relationship between subaverage Wechsler-Bellevue IQ level and GSCT performance; only Ss with an IQ of 100 or over obtained, as a group, a perfect Cube Test score. This finding indicates the existence of common factors determining performance on tests of intelligence and of abstract behavior, and suggests that the latter, like the former, may not be culture-free.

In administering the GSCT the procedure laid down by the authors (11) was followed in every detail.

Raven's Progressive Matrices. This test requires the

discovery of abstract relationships among designs of increasing complexity. It consists of symmetrical patterns, in which one element is missing. Below each pattern a set of shapes is printed, from which the one that will lead to correct completion has to be selected. This test was used during World War II with African recruits, and obtained a factor loading closely similar to that of the Block Design Test (25, p. 106), on which the GSCT is based.

The way the Progressive Matrices were applied owes its inspiration to Ombredane (22), who suggested that European tests may be more effective with Africans when administered more than once, thereby functioning as tests of educability rather than intelligence. Hence it was given on three successive weeks, with a 20-minute time limit. Otherwise Raven's instructions (23) were adhered to.

Kohs Block Design Test. The Wechsler-Bellevue version of this test ("with certain modifications in presentation suggested by Goldstein and Scheerer") had been applied by McConnell (20) to Teprehuan Indian villagers. It was administered in the manner outlined by him, and his five scoring categories were adopted.

RESULTS

Goldstein-Scheerer Cube Test

Qualitative aspects of performance. In attempting the first design most Ss clung persistently to a single block. When reminded that all four were to be used they tended to build towers like some of the "normals" reported by Goldstein and Scheerer. On the next step being presented they would reproduce the pattern correctly, but with the return to the original design Ss were apt to revert to the single block. The reason emerged quite clearly from their behavior and comments: most of them did not at first perceive the various steps as pertaining to the same task; when the original was shown again they responded once more in what they felt to be the correct way. The whole procedure was quite strange to them, so they were not surprised or annoyed on being called upon to do the same thing several times over again.⁵ With the next few designs, where more than one color was involved, the function of the several steps came to be appreciated.

The most widespread deficiency of the boys'

³ This part of the investigation was carried out jointly with Mr. A. Taylor of the Institute of Education, and the writer is indebted for permission to use some of the material for this paper.

⁴ "Should he (i.e. the 'normal') have failed on the standard design, he invariably will succeed on any one of its simplifications and also on the same standard design if presented again" (11, p. 55).

⁵ "There is definite indication of impairment of abstract behavior if a subject . . . fails again on the standard design" (11, p. 56).

⁵ Similar behavior was observed when the writer used mental mazes with African Ss. Willing to oblige the experimenter and not in a hurry to get somewhere, they may respond with the same letter for many minutes on end. One highly intelligent student (Otis Gamma IQ of 114) took over three hours to get to the end of a maze which English Ss of much lower IQ completed in less than 30 minutes.

performance, which caused them to lose more points than any other type of failure, does not seem to have been considered at all by Goldstein and Scheerer. In their reproductions Ss paid very scant attention to the spatial arrangement of the blocks on the table; designs were usually left at whatever angle they happened to be completed, so that there was nearly always some rotation. The Ss on the whole did not regard the figure-ground relationship as part of their task. This came out most clearly in cases where the multiple-choice stage was reached; when three correct models were rotated 0, 45, and 90 degrees, the answer to inquiry was almost invariably to the effect that all three were equally right.

Ignoring spatial position is something that can be fairly easily corrected by specific and emphatic instructions, as subsequent check trials with other Ss proved. Another, more fundamental aspect of the performance of the majority of the Ss was not so easily subject to improvement. It will be best to illustrate this by the actual record of the behavior of one boy.

At the outset he looked at the printed design for about 12 seconds, then turned the blocks and put them together in various ways without even once glancing at the design until 1 minute 40 seconds; during this time he once achieved the correct pattern and destroyed it again; after another 10 seconds there followed a further brief glance, and then nothing until 2 minutes 35 seconds when he examined the design for about 8 seconds; a last glance occurred just before the time was up.

It is clear from this that the procedure was not a systematic block-by-block matching, with a continuous "feedback," but rather the reproduction of an infrequently renewed memory image that was often faulty. With the designs of greater complexity a more analytical approach was frequently though not always adopted. Check trials were carried out later in an attempt to induce Ss to use the block-by-block matching method from the outset, but the success was only partial. This observation of course does not prove that it could not be done; on the contrary, the writer gained the impression that it would be possible with a more prolonged training. It is therefore doubtful whether the behavior

described is the same as Goldstein and Scheerer's "concrete dependence upon the impression of the figure, and the inability to break up the figure ideationally into single squares" (11, p. 39).

Apart from the causes of failure outlined above, several of Goldstein and Scheerer's categories were also applicable occasionally. The ones below, relating to the presented model, are listed in decreasing order of importance (numbers refer to the monograph):

3. Concrete dependence upon one global impression of the design—without articulation of parts;
4. Concrete dependence upon the "impressiveness" of certain color aspects;
1. Concrete dependence upon size.

On the whole, however, these seem to have played a relatively minor part. Failure responses to the manipulative block material occurred sometimes at the beginning, but were usually eliminated after a little handling of the blocks.

Numerical results. The mean score for the whole group was 77.6, with an *SD* of 20.3, and the range 34–115; as the possible maximum score is 120, this means that not one *S* achieved a faultless performance.

When the group is divided according to home background, an important contrast emerges. There were 17 boys from literate homes, whose mean score was 84.9; the remaining 10 from illiterate homes averaged only 65.1, a difference significant at the .02 level. The influence of environmental factors is thus clearly evident.

It would seem, therefore, that the low level of performance does not necessarily involve a lack of abstract ability, but perhaps only an inadequate development of perceptual skills. If such is the case, one would expect an improvement after a certain amount of practice with the handling of geometrical designs. On the other hand, if Africans are "incapable of sound abstraction," practice should make no difference, for Goldstein and Scheerer state that people with "impairment of abstract behavior are incapable of learning, or of acquiring the abstract attitude by learning" (11, p. 56).

Raven's Progressive Matrices

It appeared advisable to test the conclusion just outlined with a larger sample. The GSCT

TABLE 1
CUMULATIVE PERCENTAGES OF PROGRESSIVE MATRICES
SCORES
(Maximum score is 60)

Group	Scoring Levels									
	50+	45+	40+	35+	30+	25+	20+	15+	0+	
2,010 British farm-workers aged 17-19	1	8	25	48	69	85	94	98	100	
317 African school-boys (first week)	.3	1	4	17	31	46	61	74	100	
317 African school-boys (third week)	.6	9	30	46	58	64	71	82	100	

does not easily lend itself to such use, being extremely time-consuming to administer; hence the use of the Progressive Matrices Test, which taps somewhat similar abilities. As mentioned earlier the test was given on three successive weeks, with a 20-minute time limit. Table 1 shows the distribution of scores, in comparison with that of a sample of British farm workers of a somewhat higher age group. The British norms were obtained from Vernon (24), and farm workers chosen as being the group with, presumably, the least dissimilar background.

It will be seen that the scores in the third week approximate fairly closely those of the British farmworkers. The mean rise was 6.5 points or 28 per cent, a highly significant change with a critical ratio of 16.28; and contrary to the findings of Vernon and Parry (26) on a retest of British Ss, the greatest improvement took place in the middle ranges.

With a correlation of .82 between the two sets of scores the changes exhibited considerable variations, ranging from minus 24 to plus 32; uncontrolled motivational factors may have been at least partly responsible. Preliminary trials with untimed tests had shown that some of the boys became bored after the second test. That they probably lost interest to some extent even when there was a time limit is suggested by the fact that over one-quarter of the Ss declined between the second and third test, sometimes quite markedly; the extreme case of one boy may be quoted who scored 28, 35 and 4 over the series. The results would have been far more spectacular if one had selected the best scores from all three tests, but such a procedure would have capitalized on chance successes.

It may be stated with confidence, however, that the rates of improvement reported are on the conservative side.

These results support the view that increasing familiarity with the materials among which abstract relationships are to be discovered leads to a significant improvement on the part of African Ss.

Kohs Block Design Test

The proportion who came from literate homes was far smaller in this group of Ss than among those who had done the GSCT; even so the extent of literacy may be exaggerated, for it was necessary in this case to rely on indirect information obtained by questioning the boys. Yet the comparative results presented in Table 2 fail to indicate that the Ss are markedly inferior to the Tepehuan, of whom McConnell said that they function at an abstract level "almost as well as the adult population on which Wechsler standardized his intelligence test."

Thus it would seem that the African urban sample, as judged by the GSCT results, tend to be very concrete in their behavior, but a scholastically poorer semirural group could be said to be nearly as abstract as a representative Western group on the basis of their performance on the modified version of Kohs Blocks. This apparent contradiction calls for some explanation. There are three main reasons why performance on this test was relatively easier, and the handicap of the Ss from illiterate homes considerably reduced:

1. The absence of single-color designs (except among the sample items), which elimi-

TABLE 2
RANK SCORES ON KOHS BLOCKS

Rank	Adult Tepehuan Indians*			Adolescent African Schoolboys		
	Sex			Background		
	Men	Women	Total	Literate	Illiterate	Total
I	0	0	0	0	0	0
II	2	2	4	0	1	1
III	3	4	7	4	8	12
IV	7	5	12	10	8	18
V	1	0	1	0	1	1
Total	13	11	24	14	18	32

* From McConnell (20).

nated the occasion for what Goldstein and Scheerer called "concrete dependence upon size," actual or apparent.

2. The absence of diagonal designs, which meant that rotations could be ignored, as there were no specific instructions to the contrary.

3. Most important of all were the two sample designs, which enabled the experimenter to explain the task in detail to the Ss, and to ensure that they grasped the basic mode of approach before they had to start working independently.

Some of the characteristic failure responses described earlier recurred, but owing to the different conditions of administration they had less effect on the final scores. Unlike the Tepehuan sample none of the boys would have received additional credits for rapid working if Wechsler's scoring instructions (27, p. 184) had been followed.⁶ On the other hand only one single S out of the total group could be classified, according to McConnell's scheme, as being more concrete than abstract in his performance.

DISCUSSION

The series of investigations reported indicate that the problems inherent in any attempt to assess abstract behavior in different cultures are closely analogous to those encountered when one is trying to compare levels of intelligence. Neither can operate in a vacuum, and the media through which they are tested are not equally familiar to members of different cultures. It is one of the weaknesses of Goldstein and Scheerer that they do not seem to recognize this, for they write (11, p. 29): "...every normal individual—even a deaf-mute child—possesses these basic categories of like shape and color." As far as Gold Coast Africans are concerned, this is certainly not true of shape. Hebb (16) has shown that the perception of shapes is the result of a lengthy learning process. In a cultural milieu where geometrical patterns form part of a person's early experience only to a very limited extent or not at all, as is the case with most rural and

many urban Africans, the discrimination of complex shapes is likely to be a most exacting task.

Various investigations in different parts of Africa have accumulated data showing the difficulties Africans experience in manipulating spatial relations and perceiving complex shapes (2, 5, 6, 10, 21). While the possibility of some contributory genetic factor cannot be excluded a priori, the available evidence as marshalled, for instance, by Biesheuvel (4) suggests that the absence of adequate environmental support for such abilities is mainly or perhaps entirely responsible. The findings with Raven's Progressive Matrices are in line with this view.

It follows that neither the GSCT nor McConnell's adaptation of the Wechsler-Bellevue version of Kohs Block Test are suitable instruments for measuring abstract behavior in Africans, and presumably also in members of other non-Western cultures. This conclusion emerges sharply from the paradoxical results obtained in the present study with these tests.

To say this, of course, by no means implies agreement with those who hold that Africans are by nature concrete in their mentation. On the contrary, many of the African indigenous games (7, p. 110), and in the Gold Coast particularly Wari (3), require abstract ability of a high order. Thus the present writer is in agreement with part of McConnell's conclusion, though not with the method by which it was reached.

In this connection it may be observed that McConnell's belief that he discovered or proved something about "primitive thought" is based on a confusion. The misconception is best illustrated by his reference to Lévy-Bruhl, whom he describes as "contending that the thought patterns of *l'homme sauvage* are qualitatively different from those of 'modern man'..." There is an ambiguity here in the use of the term "thought pattern," which may relate to the content, a social fact, or to the process, a psychological one. In his later formulation⁷ Lévy-Bruhl made

⁶ The mean times taken for completion, converted into percentages of the maxima permitted by Wechsler, ranged from 85 per cent for the first design to 345 per cent for the last one.

⁷ One has to admit that in his major publications Lévy-Bruhl's mode of expression was sometimes ambiguous, and this has led to a widespread misunderstanding of his position. In fact, as the quotation below from a work published in 1910 indicates, he never

it quite clear that he meant the former (18, 19), and therefore no performance on a test of abstraction of the present kind, however superior, is capable of refuting him. The investigation of "primitive man's mode of thinking" would need a very different type of approach.

SUMMARY

The aim of the study was to investigate whether tests of abstract behavior like the Goldstein-Scheerer Cube Test can be validly applied to non-Western people. All Ss were adolescent boys from schools in or near Accra.

The Goldstein-Scheerer Cube Test was administered to 27 Ss. The main types of failure responses that occurred are described. It was shown that boys from literate homes performed significantly better than those from illiterate ones, evidence that Cube Test results are not unaffected by environmental influences.

In order to check this finding with a larger sample, a group test involving similar abilities was given on three successive weeks, a highly significant practice effect being obtained.

A replication of McConnell's work with the Tepehuan, using a method derived from Goldstein and Scheerer, yielded results inconsistent with those from the Cube Test. The implications of this finding were discussed, and Goldstein and Scheerer's assumptions about perception criticized in the light of research in Africa. The conclusion emerged that tests of abstract ability are no more "culture-free" than tests of intelligence.

seems to have held the view that the basic cognitive processes of "primitives" differ from those of people in more advanced societies, except in so far as they are subject to social influences.

"In this way, the exigencies of logical thought are elicited, established and then strengthened in each individual mind by the uninterrupted pressure of the social milieu, by means of language itself and through what is transmitted in the forms of language. This constitutes a heritage of which no one is deprived in our society and which no one can even have the thought of rejecting . . .

"Quite other are the conditions where the prelogical mentality obtains. No doubt it is also socially transmitted, through the intermediary of language and of concepts without which it could not find expression . . . But these concepts differ from ours and, consequently, the mental operations differ also" (17, pp. 113-4) (Writer's translation).

REFERENCES

1. ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology*. New York: Macmillan, 1949.
2. BARRÉ, Mlle. R. Psychologie. In L. Pales (Ed.), *Organisme d'enquête pour l'étude anthropologique — rapport No. 3*. Dakar: Gouvernement Général de l'Afrique Occidentale Française, 1947. Pp. 99-113.
3. BENNETT, G. T. Wari. In R. S. Rattray, *Religion and art in Ashanti*. London: Oxford Univer. Press, 1927. Pp. 382-390.
4. BIESHEUVEL, S. *African intelligence*. Johannesburg: South African Institute of Race Relations, 1943.
5. BIESHEUVEL, S. Psychological tests and their application to non-European peoples. In G. B. Jeffery (Ed.), *The yearbook of education 1949*. London: Evans, 1949. Pp. 87-117.
6. BIESHEUVEL, S. The study of African ability. Part I: The intellectual potentialities of Africans. *African Stud.*, 1952, **11**, 45-57.
7. BIESHEUVEL, S. The study of African ability. Part II: A study of some research problems. *African Stud.*, 1952, **11**, 105-117.
8. BOYD, F. A provisional quantitative scoring with preliminary norms for the Goldstein-Scheerer Cube Test. *J. clin. Psychol.*, 1949, **5**, 148-153.
9. CAROTHERS, J. C. *The African mind in health and disease*. Geneva: World Health Organization, 1953.
10. GENEVAY, M. J. La formation sociale et professionnelle des noirs. *Rev. Mens. Comité National Organ. Française*, 1949, **23**, 13-16.
11. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
12. HAVIGHURST, R. J., GUNTHER, MINNA K., & PRATT, INEZ E. Environment and the Draw-A-Man test: the performance of Indian children. *J. abnorm. soc. Psychol.*, 1946, **41**, 50-63.
13. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Goodenough scores. *Man*, 1954, **54**, 86-88.
14. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Part II, Machover scores. *Man*, 1955, **55**, 27-29.
15. HAWARD, L. C. R., & ROLAND, W. A. Some intercultural differences on the Draw-A-Man test: Part III, conclusion. *Man*, 1955, **55**, 40-42.
16. HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
17. LÉVY-BRUHL, L. *Les fonctions mentales dans les sociétés inférieures*. (9th Ed.) Paris: Presses Universitaires de France, 1951.
18. LÉVY-BRUHL, L. *Les carnets de . . .* Paris: Presses Universitaires de France, 1949.
19. LÉVY-BRUHL, L. A letter to E. E. Evans-Pritchard. *Brit. J. Sociol.*, 1952, **3**, 117-123.
20. MCCONNELL, J. Abstract behavior among the Tepehuan. *J. abnorm. soc. Psychol.*, 1954, **49**, 109-110.
21. NISSEN, H. W., MACHOVER, S. & KINDER, E. F. A study of performance tests given to a group

- of native African negro children. *Brit. J. Psychol.*, 1935, 25, 308-355.
22. OMBREDANE, A. Principes pour une étude psychologique des noirs du Congo Belge. In H. Piéron, A. Fessard, & P. Fraisse (Eds.), *Année psychol.* Paris: Presses Universitaires de France, 1951. Pp. 521-547.
23. RAVEN, J. C. *Guide to using Progressive Matrices.* London: Lewis, 1948.
24. VERNON, P. E. Occupational norms for the 20-minute Progressive Matrices Test. *Occup. Psychol.*, 1949, 23, 58-59.
25. VERNON, P. E. *The structure of human abilities.* London: Methuen, 1950.
26. VERNON, P. E., & PARRY, J. B. *Personnel selection in the British forces.* London: Univer. of London Press, 1949.
27. WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.
28. *Accelerated development plan for education, 1951.* Gold Coast Government. Accra: Government Printing Department, 1951.
29. *Annual report of the education department for the year 1952.* Gold Coast Government. Accra: Government Printing Department, 1954.
30. *Census of population 1948.* Gold Coast Government. Accra: Government Printing Department, 1950.

Received November 28, 1955.

THE MANIPULATION OF DOMINANCE IN MONKEYS WITH CONDITIONED FEAR

J. V. MURPHY AND R. E. MILLER^{1, 2}

Department of Clinical Science, School of Medicine, University of Pittsburgh

CONDITIONING as a methodological tool achieved a major advance with the advent of instrumental conditioning in that the response restrictions imposed in the classical conditioning experiments of Pavlov were largely eliminated. That is, the subject was frequently free-ranging in the experimental apparatus instead of being securely fastened in a conditioning stock. This modification sacrificed some degree of control over the conditioned response but appreciably reduced the artificial nature of the learning process. In an attempt to extend this principle, a technique for interanimal conditioning was devised (8) which relinquishes the rigid experimental control over both the stimulus and the response. With this procedure the conditioned stimulus is a monkey which is free-ranging in a relatively large compartment. Thus, the stimulus animal may be moving about, grooming himself, or performing any one of many possible behaviors during presentation to the conditioning subject. Moreover, the conditioning subject is also permitted to move freely in a relatively large compartment. In order to acquire a CR the subject, from any position in the compartment, must identify the stimulus animal regardless of its position or behavior. One might anticipate that learning in such a situation would generalize to a greater extent to other situations since varying aspects of both the stimulus and the response are incorporated in the learning process.

Mowrer (6) has advanced an interpretation of avoidance learning which utilizes the intervening variables of fear and fear reduction. Applying this theoretical interpretation to interanimal conditioning, the sight of the stimulus monkey accompanied by a painful electric shock results in an association between the stimulus animal and the pain and fear

evoked by the shock. On subsequent trials, the sight of the stimulus animal gives rise to fear in the animal undergoing conditioning. Performing the arbitrarily designated CR removes the CS (sight of the stimulus animal) and the conditioning subject is theoretically reinforced by fear reduction. A major difficulty with the fear-reduction interpretation is that the inferential evidence for the presence of fear and fear reduction is inextricably related to the conditioned response.

One advantage of the social avoidance procedure is that behavioral modifications are not necessarily restricted to the particular response or situation used to establish conditioning. If the sight of the stimulus animal gives rise to fear in the conditioned animal, the interactions of the two monkeys should be modified in a predictable fashion outside of the conditioning apparatus. Such a modification would provide independent evidence for acquired fear in avoidance learning. The dominance relationship would appear to be ideally suited for an independent assessment of "fearful" behavior. Maslow (3) has described the dominance behaviors of the Old World monkeys, including rhesus, as: "... rough, brutal and aggressive, it is of the nature of a powerful, persistent, selfish urge that expresses itself in ferocious bullying, fighting, and sexual aggression. The subordinate animal is usually afraid of the dominant animal, and is frequently completely terror-stricken and cowed" (3, p. 316). If fear is induced in a dominant monkey by the sight of a submissive partner which has been used as a CS for avoidance behavior, one would predict that the dominance relationships would be altered after the completion of conditioning.

Specifically, the present experiment was designed as an attempt to verify and extend the observation (5) that the dominance hierarchy within a group of monkeys is subject to experimental manipulation with the interanimal conditioning technique. That is, can the dominant member of a pair of monkeys be made to behave as a submissive animal follow-

¹ This investigation was supported by a research grant (M-487C) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service; and in part by the Foundations' Fund for Research in Psychiatry.

² The authors are indebted to Dr. I. Arthur Mirsky for suggestions and criticisms.

ing avoidance conditioning in which the submissive member of the pair is the conditioned stimulus? An incidental aim of the present study was to obtain evidence on the presence of acquired fear in avoidance learning which is independent of the conditioning situation.

METHOD

Subjects. Ten experimentally naive male rhesus monkeys were used as Ss. The weight-range in the group was 9½ to 18 lbs. when the study was undertaken and 12 to 25 lbs. at the finish.

Apparatus. The conditioning apparatus (8, Fig. 1) consisted of a large rectangular box divided into two compartments by a specially constructed one-way vision screen which made it impossible to view the stimulus compartment from the conditioning compartment. A trial was instigated when the experimenter altered the illumination differential in such a manner as to make the stimulus compartment clearly visible from the conditioning compartment. The shock-unconditioned stimulus was automatically controlled by an Agastat relay. The response bar, mounted on the one-way vision screen, was modified slightly from that reported previously (7, 8). The essential modification was that the bar for the present experiment required a downward thrust instead of a pulling response. This modification was made primarily for additional studies and was only incidental to the present experiment. Dominance tests were conducted in the Wisconsin General Test Apparatus (WGTA) described by Harlow and Bromer (2).

Procedure. The Ss were hand-tamed daily in the WGTA for eight weeks prior to the beginning of the investigation. The daily taming trials were arbitrarily terminated after 25 raisins had been taken by the animal or after 15 minutes. Hand-taming was continued until such time as all Ss were responding satisfactorily in the test situation. The animals were all housed separately and were contained in closed-side cages to eliminate interactions with other members of the group outside of the experimental situation.

Following adaptation training the dominance structure of the entire group was determined. The dominance relationships were determined by placing a pair of animals in the restraining cage of the WGTA. The test was instigated by prefeeding each monkey with a raisin. On the succeeding ten trials a single raisin was placed on the moving tray of the WGTA and the attention of the animals drawn to it by tapping. The tray was then moved within reach of the animals and the experimenter recorded which animal obtained the food. The animal obtaining the greater number of raisins was designated as the dominant animal. None of the moving screens on the WGTA was used throughout the experiment, i.e., the experimenter was in full view throughout the test. The dominance testing procedure has been more fully described elsewhere (4). Each subject was paired with all other monkeys in the group. Tests were conducted at the rate of 10 pairs a day for four days with five pairings on the fifth day. The dominance structure was retested for the entire group starting two weeks after the beginning of the first test.

Two weeks after the second group-dominance test was started, the animals were ranked on the basis of the two group tests. Adjacent animals were then tested in the dominance situation for five consecutive days. Thus, on each of the five pairs there were a total of seven dominance tests. The monkey which obtained the larger number of raisins on the greater number of the seven tests was designated as the dominant member of the pair.

In the interanimal conditioning phase the submissive animal was the conditioned stimulus for conditioning an avoidance response in the dominant animal. During conditioning the dominant member of each pair was placed in the conditioning compartment while the submissive member of the pair was placed in the stimulus compartment. On a conditioning trial the experimenter switched the lighting arrangement so that the stimulus animal was clearly visible from the conditioning compartment. Five seconds following the exposure of the stimulus animal, a 50-volt electric shock was automatically delivered through the grid floor to the animal being conditioned. Pressing the response bar enabled the animal to escape the shock or to avoid it if the response occurred within a five-second interval. The instrumental response also removed the stimulus monkey from view, i.e., response-terminated stimuli were employed. Twenty trials a day were administered with an average of 60 seconds between trials. Ten conditioning and 10 control trials were presented in a randomly interspersed pattern each day. On control trials the conditions were identical with the exceptions that the stimulus animal was removed from the apparatus and the trials were terminated manually after five seconds. No shock was present on the control trials. Conditioning was continued to a criterion of 90 per cent conditioned responses for three successive days with not more than 10 per cent responding on control trials for the same three days.

Twenty-four hours after the conditioning subject in each pair attained the conditioning criterion, dominance retesting in the WGTA was begun. The dominance status of the pair of monkeys was determined on five successive days. The data from the dominance tests following completion of conditioning appeared to be confounded somewhat by relatively poor adaptation to the WGTA. This difficulty was apparently due to the prolonged absence from the dominance test apparatus while conditioning was under way. In order to eliminate this source of contamination the animals were reconditioned after a 40-day interval. The original procedure was followed with the additional control that for 10 days preceding reconditioning and throughout reconditioning each animal was hand-tamed daily in the WGTA. This taming procedure consisted of presenting 12 single raisins on the moving tray of the WGTA. Each animal was tamed alone during this period. After attaining the reconditioning criterion, the five dominance tests for each pair were repeated. Two weeks after the five pairs of animals completed the experimental regimen, the dominance structure for the entire group was determined, i.e., every animal was tested with every other animal. The group dominance structure was reassessed after an interval of one week. One animal was excluded from the study prior to this

group testing since it appeared to be in ill health. This animal died some three months later.

On the basis of the interim group test, three additional pairs of monkeys were selected for study. Again an attempt was made to work with pairs of animals which were as nearly adjacent as possible in the dominance hierarchy. One animal selected to be conditioned to the sight of another monkey had previously been conditioned with another animal as the stimulus in an earlier pairing. The remaining two conditioning Ss had served as stimulus animals in the first five pairings.

These three pairs of animals were given identical treatment to that described for the original pairings. That is, the dominance of each pair was tested for five successive days, conditioning was completed to criterion, and five paired tests were undertaken. Conditioning was then repeated as previously described with the inclusion of hand-taming trials during reconditioning, etc.

RESULTS

The dominance data for the submissive animal in each pair on the seven pretests (two group tests and five paired tests) are summarized in Table 1 along with the number of tests

TABLE 1
PERCENTAGE OF TESTS DOMINATED BY THE ORIGINALLY
SUBMISSIVE MEMBER OF EACH PAIR

Pair	1	2	3	4	5	6	7	8
Pretests	0	7	0	29	43	36	0	36
Posttests	90	0	60	100	60	100	80	60

dominated by the submissive member on the five dominance measures which followed the completion of reconditioning. Data from the five dominance tests following original conditioning will not be reported. These data show the same trends but are less definitive.

The significance of changes in dominance following completion of conditioning was determined by computing a *t* with the difference technique (1) between the mean number of tests dominated by the submissive animal before and after conditioning. This comparison resulted in a *t* of 4.17 which with seven degrees of freedom is significant at the .01 level.

The acquisition of the original avoidance response was found to be very similar to that described in a previous study on interanimal conditioning (8). The number of conditioning trials required to attain the criterion ranged from 100 to 290.

DISCUSSION

The present study was designed in an attempt to verify and extend the observation

that dominance relationships may be experimentally manipulated with the interanimal avoidance conditioning technique (5). After the dominant member of each of the eight pairs of animals had been conditioned with the submissive partner as the conditioned stimulus, the dominance relationship was reversed as predicted in seven of the eight pairings.

It is important to note that the conditioning tests and the dominance tests were conducted in separate and distinct situations. In addition to many distinctive features of the test situations, the experimental manipulations were conducted in separate rooms. In this case the learning in the conditioning situation generalized to a new situation and modified the subsequent behavior of the conditioned animals. This observation indicates that conditioning principles are not necessarily restricted to the artificial nature of soundproofed rooms and "pure" stimuli.

Not only did the CS acquire a new significance for the conditioned animal as is true of conditioning studies generally, but the CS in this instance acquired a meaning in an opposite direction to that existing prior to conditioning, i.e., a monkey which previously elicited dominant behavior now elicited submissive behavior.

This dramatic alteration in dominance behavior is somewhat surprising in that with the interanimal conditioning technique, the experimenter is only manipulating one aspect of a diadic relationship, i.e., the dominant animal presumably learns to become fearful at the sight of the submissive monkey. One might predict an enhanced effect if the submissive animal were taught in some manner to be nonfearful of the previously dominant animal. On several occasions it appeared clear that the previously dominant partner was not prepared to do combat to obtain the piece of preferred food; however the previously submissive animal in such pairings was also hesitant to take the food-getting initiative.

Although dominance determinations on the entire group of monkeys were important in determining the relative status of an animal in the group, it appeared desirable to work with single pairs of animals when the independent variable was administered. This modification in procedure from a previous study (5) made it possible to continue con-

ditioning to the same criterion for all animals. In addition, the interactions between dominance relationships which occur during group testing were eliminated by administering the critical dominance tests to pairs of animals.

The observed modifications in interindividual transactions conform very nicely with that suggested by the interpretation of avoidance conditioning which emphasizes the presence of fear as an intervening variable. The modification of dominance provides behavioral evidence for the presence of fear in avoidance learning which is independent of the conditioning situation. Recently completed experiments on interanimal conditioning in which the extinguished response in the previously conditioned animal is reinstated at a high level by observing the stimulus animal undergoing painful shock stimulation will make it possible to establish more definitive evidence for the presence of fear in avoidance conditioning and its communication between animals.

In the dominance tests following completion of interanimal conditioning the animals were carefully observed to note any deviation from the typical behavior of monkeys in this test situation. No differences could be distinguished, i.e., there was no apparent "bar-searching" activity, etc.

Since the modification in dominance following interanimal conditioning was attributed

to the originally dominant animal learning to become fearful at the sight of the previously submissive animal, it was a logical extension to predict that as the conditioning was extinguished the dominance behavior would again revert to the original relationship. Due to the experimental design, only the last three pairs of animals could be used in determining this relationship; consequently, the data are more suggestive than definitive. In this phase of the experiment the extinction trials, following an additional reconditioning, were run each morning (except Sunday) and the dominance determinations were taken each afternoon.

The dominance pattern in all three pairs was found to be reversed to the original relationship as extinction training was continued. The number of conditioned responses during extinction and the number of raisins in the dominance test for a typical animal previously made dominant during the conditioning phase are presented in Fig. 1. Due to the startling correspondence between day-to-day fluctuations in the two experimental situations, the precaution was instigated of having one experimenter run the conditioning trials while the other conducted the dominance tests without any information about the conditioning data for that day; however, the basic relationship continued as before.

The relationship shown in Fig. 1 has interesting implications for the theoretical inter-

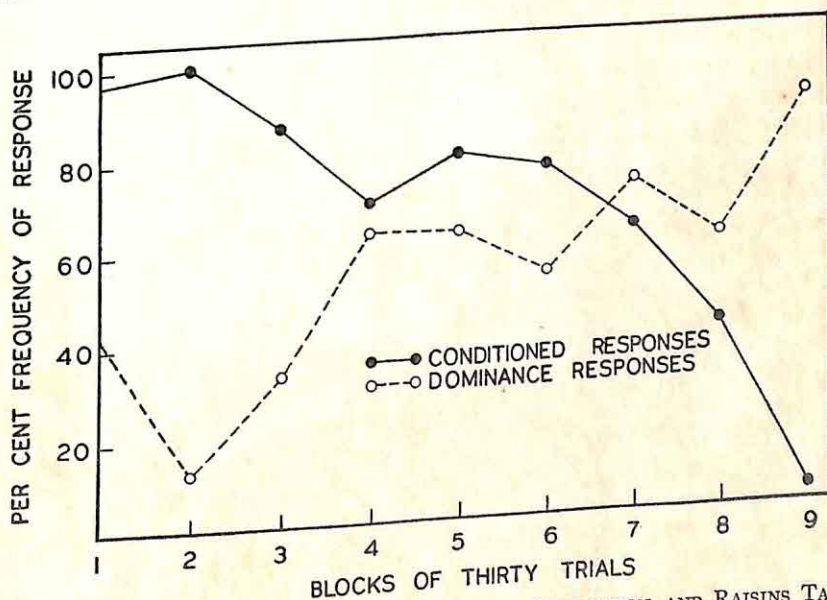


FIG. 1. CONDITIONED AVOIDANCE RESPONSES THROUGHOUT EXTINCTION AND RAISINS TAKEN IN THE DOMINANCE TEST BY A TYPICAL MONKEY

pretation of the extinction process of an avoidance response. In typical avoidance studies it is unknown whether the decrease in responding during extinction is attributable to extinction of the motor response, the fear response, or a combination of the two. The postulation (9) that the fear attendant to CS presentation decreases on the nonreinforced presentations during extinction with the CR subsequently diminishing in frequency is supported by the relationship depicted in Fig. 1. In this case an independent measure of the capacity of the CS to elicit fear at varying stages throughout extinction was available. Since the measure of fear in the social situation (number of raisins taken in the dominance test) increased at approximately the same rate that conditioned avoidance responses decreased, it suggests the inference that extinction of the bar-pulling response is primarily attributable to decreased motivation (fear) and not to fatigue accompanying the conditioned response, as many theorists would lead us to believe.

SUMMARY

Two group dominance tests were conducted on ten rhesus monkeys. On the basis of these tests five pairs of animals adjacent, or nearly adjacent, in the hierarchy were given an additional five dominance determinations. The animal in each pair which received the greater number of raisins in each of the seven tests between the two animals was designated as dominant. This animal in each pair was subjected to avoidance conditioning with his submissive partner as the conditioned stimulus.

After attaining a conditioning criterion, each pair of monkeys was given an additional five dominance tests.

After completing this phase two additional group-dominance tests were completed. An additional three pairs were selected as before and subjected to the same procedure. The dominance status was found to be significantly reversed following the completion of conditioning. It was suggested that this observation provides behavioral evidence for the presence of fear in avoidance conditioning which is independent of the conditioning situation.

REFERENCES

1. EDWARDS, A. L. *Statistical analysis*. New York: Rinehart, 1946.
2. HARLOW, H. F., & BROMER, J. A. A test apparatus for monkeys. *Psychol. Rec.*, 1938, 2, 434-436.
3. MASLOW, A. H. Dominance-quality and social behavior in infra-human primates. *J. soc. Psychol.*, 1940, 11, 313-324.
4. MILLER, R. E., & MURPHY, J. V. Social interactions of rhesus monkeys: I. An examination of the stability of dominance relationships. *J. soc. Psychol.*, in press.
5. MILLER, R. E., MURPHY, J. V., & MIRSKY, I. A. The modification of social dominance in a group of monkeys by interanimal conditioning. *J. comp. physiol. Psychol.*, 1955, 48, 392-396.
6. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
7. MURPHY, J. V., & MILLER, R. E. A flexible general test apparatus for the monkey. *J. gen. Psychol.*, in press.
8. MURPHY, J. V., MILLER, R. E., & MIRSKY, I. A. Interanimal conditioning in the monkey. *J. comp. physiol. Psychol.*, 1955, 48, 211-214.
9. MURPHY, J. V., MILLER, R. E., & FINOCCHIO, D. V. Spontaneous recovery of an avoidance response over an extended time interval in the monkey. *J. genet. Psychol.*, in press.

Received May 11, 1956. Prior publication.

CRITIQUE AND NOTES

AN INVESTIGATION OF PHONETIC SYMBOLISM¹

IRVING MALTZMAN, LLOYD MORRISETT, JR.,² AND LLOYD O. BROOKS

University of California, Los Angeles

DURING the course of several years we have conducted a series of studies that are directly related to the experiments on "phonetic symbolism in natural languages" recently reported by Brown, Black, and Horowitz (1). They found, in accord with the work of earlier investigators, that Ss could match foreign words with their English equivalents significantly beyond chance expectancy. The languages employed were Chinese, Czech, and Hindi. They interpreted their results in terms of a universal phonetic symbolism common to natural languages. This universal phonetic symbolism is based upon the hypothesis of physiognomic language which assumes that there are unlearned interconnected sensory systems underlying sounds and meanings, an application of the gestalt principle of organization that communication may occur among different trace systems that are organized in a similar fashion. Foreign words can be matched with their English equivalents because they are both related to similarly organized neural trace systems.

The principle of gestalt organization was invoked by Tsuru and Fries (2) to account for their finding that Ss could select the Japanese equivalents of English words with a high degree of success. The results of the present experiments confirm and extend the empirical results of these previous investigators by obtaining similar results with a foreign language not previously employed. However, it is believed that certain of our results cast doubt upon the foregoing interpretation of the phenomena.

The present experiments were all conducted in group form. The Ss were students in the senior author's psychology classes. At the beginning of a class hour, E would distribute questionnaires to the students and make a brief announcement requesting their cooperation in the investigation. The questionnaire contained 25 stimulus words and 25 pairs of response words, one pair for each stimulus. A sample of the instructions are as follows:

Below is a list of Japanese and English words. For each Japanese word I want you to select the corresponding English word. Of course you do not know the

meaning of the Japanese words. But try to pick the English word that seems to best fit the Japanese word.

In order for you to be successful, it is necessary that you avoid artificial terminological schemes such as associating sounds, letters, length of words, etc.

Underline the word that you think is correct. Please do not ask any questions or call out until the papers are collected.

No further instructions or illustrations concerning the pronunciation of the foreign words were given.

The Ss were allowed to work on their answers for as long as they wished. The time required was not more than 10 minutes. The Ss were also asked to indicate whether they had any acquaintance with the foreign language(s) in question. The results of seven Ss were discarded from the various experiments. Six Ss were of Japanese descent and one had spent several years in Japan in the Army of Occupation.

The first experiment was designed to determine whether the order of the English and the foreign words, Japanese in this case, and the nature of the student sample would influence the frequency of choice of the equivalents. In the Japanese-English mode of presentation, the 25 Japanese words were the stimuli and English words were the response pairs. One member of each pair was the equivalent of the Japanese stimulus word. In the English-Japanese mode of presentation, 25 English words were the stimuli and Japanese words formed the response pairs. The first and second columns in Table 1 contain the English words and the Japanese equivalents employed in this and subsequent experiments. All of these words were obtained from the article by Tsuru and Fries (2). The Japanese stimulus words for the first list were randomly selected from each pair of antonyms in their list. The order of the correct and incorrect member of each response pair was varied randomly from pair to pair. The English stimulus words in the second list were the equivalents of the Japanese words used in the first list.

A total of 65 Ss served in this experiment. There were 25 undergraduate and 40 graduate Ss in the Japanese-English condition, and 16 undergraduate and 19 graduate Ss in the English-Japanese condition. Mood's nonparametric median test of the number of errors committed by the four different groups was not significant, $\chi^2 = 3.29$, $df = 3$, $p > .30$. The groups were therefore combined.

¹ This study was aided by a grant from the Committee on Research, Graduate Council, University of California.

² Now at Yale University.

TABLE 1
ENGLISH WORDS AND THEIR JAPANESE AND CROATIAN EQUIVALENTS

English	Japanese	Croatian	English	Japanese	Croatian
1. bird	tori	ptica	13. high	takai	visok
worm	mushi	crv	low	hikui	nizak
2. red	aka	crven	14. kite	take	zmaj
green	midori	zelen	boat	hune	čamac
3. peace	heiwa	mir	15. old	oitaru	star
war	tatakai	rat	young	wakai	mlad
4. sweet	amai	sladak	16. hot	atsui	topal
bitter	karai	gorak	cold	samui	hladan
5. fast	hayai	brz	17. are	aru	jesu
slow	osoi	spor	are not	nai	nisu
6. white	shire	bijel	18. blue	ao	plav
black	kuro	crn	yellow	kii	žut
7. square	shikaku	četvorina	19. thick	atsui	debel
circle	maru	krug	thin	usui	tanak
8. good	yoi	dobar	20. big	ookii	velik
evil	warui	zao	small	chiisai	malen
9. praise	homeru	hvaliti	21. clear	sunda	jasan
deprecate	kenasu	kuditi	muddy	nigotta	mutan
10. far	tooi	dalek	22. enemy	mikata	dušman
near	chikai	bliz	friend	teki	prijatelj
11. soft	yawai	mekan	23. crooked	magatta	savijen
hard	katai	tvrd	straight	massugu	ravan
12. smart	riko	pametn	24. right	tadashii	prav
dull	baka	glup	wrong	ayamereru	kriv
			25. sharp	togatta	oštar
			dull	nibui	tup

The mean number of errors obtained by these Ss was 11.11, $\sigma = 2.67$, where 12.50 errors would be expected by chance. The obtained divergence from chance is highly significant, $CR = 4.21$, $p = < .0001$.

The results of this experiment confirm the findings of Tsuru and Fries (2) and indicate that the order of presentation of the English and Japanese equivalents does not have a significant effect. Academic level, likewise, did not influence significantly the frequency of choice of the correct equivalent.

As Brown *et al.* (1) point out, however, the lists employed in this first experiment were constructed by Tsuru who was familiar with both Japanese and English, thereby providing a possible source of bias. This difficulty is avoided in the next experiment where Croatian equivalents of the English stimulus words of the English-Japanese list were employed.³ Forty-eight Ss received a list of 25 Croatian stimulus words with adjacent pairs of English response words, one of which was the equivalent. The Croatian words used in this and in subsequent studies are given in the third column of Table 1.

The obtained mean number of errors in this

³ We are indebted to Mr. Lotar Zahradka for providing the Croatian equivalents.

experiment was 11.31, $\sigma = 1.97$, $CR = 4.25$, $p = < .0001$. These results clearly indicate that the English equivalents of Croatian words can be selected with a frequency significantly beyond chance expectancy, and quite as effectively as the selection of English and Japanese equivalents.

The results of these two experiments are in accord with the empirical findings of Brown, *et al.*, (1), and Tsuru and Fries (2), and appear to support their hypotheses of phonetic symbolism and gestalt organization of trace systems.

If the above theoretical interpretation is correct, however, the presence of an English word in a stimulus-response pair is not essential for a correct match. The Ss given pairs of Japanese and Croatian words should be able to select the appropriate equivalents as readily as in the case with English and foreign words, since the interconnections or communication pathways among the sensory systems related to the equivalents are unlearned and universal.

In the third experiment, therefore, Croatian words were used as the stimulus words with Japanese response words, one member of each pair being an equivalent of the Croatian word. Twenty-nine Ss served in this experiment. The mean number of errors obtained was 12.28, $\sigma = 2.57$, $CR = .46$, $p = > .60$.

In order to make certain that the order of pres-

entation was not a variable responsible for the lack of significance in this experiment, a fourth experiment was conducted. Forty Ss were given Japanese stimulus words with Croatian response words. The results were the same as in the previous experiment. A mean of 12.15 errors was obtained, $\sigma = 2.76$, $CR = .80$, $p = > .40$.

The results of the last two experiments therefore appear to be contrary to the gestalt principle of organization and the notion of a universal phonetic symbolism. It seems that for matching to be successful by English-speaking subjects, one member of the match must be an English word. This implies that the ability to match foreign equivalents is based upon complex kinds of learning, involving, we would surmise, mediated generalization.

SUMMARY AND CONCLUSIONS

Four experiments were conducted investigating the accuracy with which Ss can correctly match English words and foreign equivalents. The first experiment conducted showed that Ss could match

English and Japanese equivalents at a level beyond chance expectancy, and that there was no significant difference between undergraduate and graduate Ss in this respect. A second experiment showed that English and Croatian words could be matched at a level beyond chance expectancy. However, two additional experiments showed that the correct equivalents among Japanese and Croatian word pairs were not matched at greater than a chance level. The latter finding was interpreted as being contrary to the hypothesis of the gestalt organization of trace systems, and the related hypotheses of phonetic symbolism and physiognomic language.

REFERENCES

1. BROWN, R. W., BLACK, A. H., & HOROWITZ, A. E. Phonetic symbolism in natural languages. *J. abnorm. soc. Psychol.*, 1955, 50, 388-393.
2. TSURU, S., & FRIES, H. S. A problem in meaning. *J. gen. Psychol.*, 1933, 8, 281-284.

Received August 29, 1955.

PSYCHOTHERAPY AND THE RECOVERY FROM NEUROSIS

DALBIR BINDRA

McGill University

BOTH practically and theoretically, it is important to find out whether or to what extent psychotherapy is effective in the treatment of various behavioral disorders. This paper considers from a critical and broad viewpoint the problem of the efficacy of psychotherapy in the treatment of neurosis. It separates and clarifies the specific issues involved, and attempts to show what questions can or cannot be answered on the basis of available evidence.

PROBLEM

Basically, the question is whether or not the proportion of recoveries is greater among neurotics who undergo psychotherapy than among neurotics who do not receive such therapy. Meehl, in his recent review of the literature on this subject (5), notes the lack of the type of controlled studies which would unequivocally answer this question. In the absence of adequate experimental evidence on this question, psychologists have tried to argue (e.g., 2, 3, 8) for or against the efficacy of psychotherapy on the basis of indirect evidence obtained from improperly controlled studies. In general these arguments reach one of two conclusions: "No one has yet demonstrated that psychotherapy is effective in treating neurosis," or "No

one has yet demonstrated that psychotherapy is not effective in treating neurosis." Clearly, this question is amenable to experimental attack. With two groups of patients equated in terms of symptoms, severity of neurosis, motivation for therapy, socioeconomic status, intelligence, and the like, only one of the groups must receive psychotherapy. A comparison of the proportion of recoveries in the two groups will then indicate whether psychotherapy is effective.

Unfortunately, discussions of this rather straightforward empirical question can easily get complicated by certain theoretical preconceptions. In the thinking of many psychologists, the problem of the efficacy of psychotherapy is intimately tied up with the problem of the etiology of neurosis. The notion that life experiences play a dominant and crucial role in the onset of neurotic ailments is generally considered to support the view that psychotherapy can cure these ailments. And the view that neuroses arise primarily from organic or nonpsychological factors is considered to favor the view that psychotherapy is ineffective. No one has explicitly stated and defended this type of argument, but it is implicit in most reputable psychological discussions (e.g., 9). There is little justification, however, for this contention

that etiology and therapy bear such a direct and clear-cut relation to each other. The only effective treatment of a disorder of organic origin (like cerebral palsy) may consist of some form of training or psychotherapy; on the other hand, a psychogenic disorder (like some insomnias) may be most effectively treated by medication. The mere knowledge of the etiology in these cases does not tell whether a particular treatment would be effective. Similarly, whether neurotic ailments result from psychological conflicts, environmental stresses, or some subtle chemical factors in the blood is quite irrelevant in determining the efficacy of psychotherapy in their treatment. I do not mean to imply that neurotic disorders can be identified in a way that excludes etiological concerns. Rather, the point here is simply that the effectiveness of psychotherapy is an issue that can be and should be decided without linking it with any specific conjecture about the etiology of neurotic ailments.

Clarity and explicitness are also necessary in defining neurosis. In the present discussion, *neurosis* or neurotic disorder refers to cases of persistent and gross maladjustments involving some definite, palpable, behavioral peculiarities (or symptoms) such as anorexia, phobias, compulsive acts, demonstrable anxiety attacks, hysterical blindness, impotence, amnesias, and the like. It should be noted that this rough and arbitrary definition excludes minor maladjustments defined with reference only to subjective states such as "unhappiness," "anxiety," and "tension." It is true that "anxiety," "unhappiness," and "tension" are often (though not always) reported by neurotics, but they are also frequently reported by normals. Since these states are not a *sine qua non* of persistent and gross maladjustments unless accompanied by definite symptoms of the type listed above, they are excluded from the present definition of neurosis. Some psychologists believe that both neurosis and these minor maladjustments arise from the same type of etiological factors. Even if this view is correct, it does not necessarily imply that an effective treatment of minor maladjustments would also be effective in cases of neurosis, or vice versa. The concern here is primarily with recovery from neurosis.

A GENERAL FORMULATION

Recovery from neurosis means nothing more than a kind of behavioral change, a change in those aspects of the patient's behavior that put him in the category of neurotics. Thus, the question of the effectiveness of psychotherapy is really a special case of the general problem of the extent to which psychotherapy can produce *personality change*. A clear formulation of this question requires a closer analysis of the concepts of "per-

sonality change" and "psychotherapy." Precise descriptions are needed of (a) the personality variables along which change is presumed to occur, and (b) the psychotherapeutic processes which are presumed to effect the change. The effect of a specified therapeutic process on a specified personality variable can then be investigated.

Personality variables can be specified easily enough. For present purposes, they can arbitrarily be categorized into five rough categories: (a) *cognitive variables*, such as general information, reasoning ability, and memory; (b) *attitude variables*, such as attitudes toward in- and outgroups, attitude toward oneself, and likes and dislikes; (c) *need variables*, such as introversion, super-ego, anxiety, and aggression; (d) *specific neurotic symptoms*, behavioral characteristics, such as hysterical or compulsive symptoms, which are used in defining persons as neurotics; (e) *personality variables (if any) that are causally related to neurotic ailments*. Intelligence tests, scales of attitudes, and "personality tests" could be used to measure changes along the first three types of personality variables. The last two categories are not at the same level of discourse as the first three, nor have they been adequately identified so far, but they are listed here to focus attention on the type of variables that are of direct concern in this discussion. Accepting this admittedly arbitrary categorization, the problem of the effect of psychotherapy on personality variables resolves itself into five separate questions. Does psychotherapy have any effect on (a) cognitive variables, (b) attitude variables, (c) need variables, (d) specific neurotic symptoms, and (e) personality variables (if any) that cause neurotic symptoms?

To answer these questions, it is necessary to specify the exact nature of psychotherapy.¹ Unfortunately, it is not possible to do so. Psychotherapy still has its mysterious aspects, and present knowledge is far from unraveling the essential nature of all that takes place in the therapist's office. But from all accounts, it appears that the therapist provides the patient with a friendly, permissive, and uncritical atmosphere, and then systematically makes use of "interpretation," suggestion, and catharsis. It is not clearly understood, however, exactly how these various processes operate in the psychotherapeutic situation. Ideally, empirical questions should be formulated in terms

¹ This conception excludes the *incidental* "psychotherapy" that is sometimes said to have taken place when a neurotic gets some attention from a nurse or when he confides his innermost thoughts to his bartender, priest, or barber. Whether or not incidental "psychotherapy" has some of the same ingredients as psychotherapy proper (given by a qualified psychiatrist or clinical psychologist) is impossible to say until more is known about what is involved in the latter.

of specific psychotherapeutic processes. Thus, one could ask such questions as "Can suggestion change attitudes and needs?" or "Can interpretive analysis affect compulsive or hysterical symptoms?" or "Can catharsis relieve functional impotence?" Such questions, however, cannot be answered unless psychotherapy is analyzed into its component processes rather than conceived as a unitary whole. The only point here is that if knowledge of exactly how psychotherapy produces whatever effects it does produce is to be gained, then an analysis of this type must be undertaken. From the point of view of the therapist, psychotherapy may well be considered as a unitary healing device, but for the researcher it is an analyzable complex of psychological processes. For the present, however, one is forced to discuss the issues without benefit of this kind of analysis. With this handicap in mind, questions concerning the effects of psychotherapy on the various categories of personality variables may now be considered.

EFFECTS OF PSYCHOTHERAPY

Inasmuch as psychotherapy involves prolonged social interaction between the therapist and the patient, it is to be expected that the psychotherapeutic situation will produce change along at least some personality dimensions. It is known that interpersonal relations can affect behavior. Interests and tastes change through association with friends. Gestures of immigrants begin to conform to the typical gestures of their adopted communities (1). Exposure to conflicting social influences may contribute to behavioral disturbances in juveniles (4). Attitudes can be changed, one way or the other, simply by showing a short movie to audiences (6). In view of the abundance of naturalistic and experimental evidence of this kind, it would be strange, indeed unbelievable, if the long and intimate social interaction between the therapist and his patient did not produce some change in the patient (and, indeed, in the therapist himself). The patient is likely to learn something about the work of the therapist, to increase his vocabulary, and to change his attitudes. Since he is the focus of attention in the therapeutic situation, the patient is particularly likely to learn more about himself and to evaluate himself rather differently. Also, the change from a nagging wife, a rude boss, and critical friends to the understanding and permissive therapist is likely to facilitate these new self-evaluations, and they may be more in keeping with the therapist's direct or indirect suggestions regarding what a normal person thinks of himself.

The recent studies of Rogers, Dymond, and their collaborators (7) present direct evidence in

support of these statements. They administered a number of personality tests to their patients before, during, and after psychotherapy. Comparable control data were obtained from normal subjects, who were not given therapy, and from patients who had to wait for therapy. The most impressive finding was that psychotherapy produced a significant change in self-perception or in the perception of the self-ideal. It was also found that responses of patients to the Thematic Apperception Test changed as a result of psychotherapy in the direction of greater judged personality integration and adjustment. The results on needs and changes in behavior in everyday life situations were equivocal, and no change in attitudes toward others could be attributed to the treatment. No data are presented on the changes in specific neurotic symptoms. It seems appropriate to conclude that psychotherapy has some effect on the client's attitudes toward himself. That specific psychotherapy (as well as informal interpersonal relations generally) can produce behavioral changes does not, of course, diminish the need for analyzing "psychotherapy" into its component processes, as suggested in the last section. If one could identify the specific effective agents in the psychotherapeutic situation, it would be a step toward more efficient therapeutic procedures.

These demonstrations by Rogers, Dymond, and collaborators, important as they are, do not, of course, tell anything about the efficacy of psychotherapy in the cure of neurosis. These workers have shown that the attitudes of their patients changed, and that, *on the basis of test data*, they were said to become more "mature" and "better adjusted." But this does not indicate whether psychotherapy was effective in eradicating the specific neurotic symptoms of their clients. Even if psychotherapy demonstrably alleviates some minor maladjustments involving attitudes of different kinds, the issue of the effectiveness of psychotherapy in the treatment of the basic, gross, and persistent maladjustment ("primary neurotic symptoms") still remains unanswered. A neurotic with hysterical or compulsive symptoms may well become better adjusted as a result of psychotherapy without showing any change in his primary neurotic symptoms. Even a person who loses a leg may benefit from psychotherapy inasmuch as his life-goals and attitudes may change in a way that would be more consistent with his handicap, and in this sense he may become more mature, better adjusted, and happier. However, in order to demonstrate that psychotherapy is effective in curing neurosis, it must show that it diminishes or eradicates the symptoms that constitute the neurosis, quite apart from making the patient better adjusted, mature,

or happier. Such a demonstration has not yet been made. Perhaps future research will show that certain psychotherapeutic processes can be effective in the treatment of certain types of neurotic maladjustments.²

Even if it were demonstrated that psychotherapy can relieve neurotic ailments, there would still be no answer to the final question. Can psychotherapy produce a change in those processes or personality variables (if any) that are causally connected with neurosis? Evidence showing psychotherapy to be effective in eradicating neurotic symptoms would not indicate anything about the mechanism by which psychotherapy operated. In this connection, two types of treatments should be distinguished. One kind involves directly undoing the disease process analogous to treating streptococcal infections with penicillin. The other type of treatment is indirect and is aimed at something other than the disease process itself; an analogy is treating typhoid or tuberculosis with rest. In textbooks on psychiatry and abnormal psychology, it is generally implied that the effectiveness of psychotherapy results from directly undoing (through relearning) the processes that produce neurosis. However, there is no evidence for this belief, and there cannot be until we know the exact basis of neurosis. It is quite conceivable that, were psychotherapy shown to be effective in treating neurosis, its effectiveness could result from some indirect effects that are quite unrelated to the causes of neurosis.

In summary, available evidence suggests that psychotherapy can be effective in alleviating

² Whether the recovery from neurosis could be more effectively produced by methods other than psychotherapy is an important question, but one that need not be raised in the present discussion.

minor maladjustments characterized by such subjective states as "unhappiness," "anxiety," and "tension" that often accompany neurosis. But the evidence does not yet support the view that psychotherapy is effective in relieving those gross and persistent maladjustments (neuroses) that are characterized by definite, palpable symptoms. Whether it has any effect on these is a question which only future research can settle.

REFERENCES

1. EFRON, D., & FOLEY, J. P. Gestural behavior and social setting. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 33-40.
2. EYSENCK, H. J. The effects of psychotherapy: an evaluation. *J. consult. Psychol.*, 1952, 16, 319-324.
3. EYSENCK, H. J. The effects of psychotherapy: a reply. *J. abnorm. soc. Psychol.*, 1955, 50, 147-148.
4. HUNT, J. McV. An instance of the social origin of conflicts resulting in psychoses. In C. Kluckhohn & H. A. Murray (Eds.), *Personality in nature, society and culture*. New York: Knopf, 1948. Pp. 367-374.
5. MEEHL, P. E. Psychotherapy. *Annu. Rev. Psychol.*, 1955, 6, 357-378.
6. PETERSON, R. C., & THURSTONE, L. L. *Motion pictures and the social attitudes of children*. New York: Macmillan, 1933.
7. ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univ. of Chicago Press, 1954.
8. ROSENZWEIG, S. A transvaluation of psychotherapy—a reply to Hans Eysenck. *J. abnorm. soc. Psychol.*, 1954, 49, 298-304.
9. WHITE, R. W. *The abnormal personality*. New York: Ronald, 1948.

Received September 3, 1955.

A SHORT FORM OF WITKIN'S EMBEDDED-FIGURES TEST¹

DOUGLAS N. JACKSON

The Menninger Foundation

SINCE Gottschaldt (2) studied the role of past experience in perception by providing differential experience with simple forms and later determining how well they were masked when contained in a more complex pattern, various modifications (e.g., 3, 4) of the original patterns have been developed to measure individual differences in perceiving. Recently Witkin (5) has

developed an embedded-figures test, based on the Gottschaldt figures, but more difficult than the latter due to the addition of color to the complex patterns. In an extensive study of perception-personality relationships by Witkin and his colleagues (6) this test proved to correlate significantly with measures of orientation to the upright, as well as with a host of personality variables. Sex differences in scores on this test, favoring males, have consistently emerged.

Of the perceptual variables used in the Witkin

¹ This research was completed while the author was a Public Health Service Postdoctorate Research Fellow at The Menninger Foundation.

studies, the embedded-figures test is the only one which does not require expensive apparatus, and hence will probably be the one most frequently used by other research workers. A major difficulty exists in administering this test. Since the score is the time required to extract simple figures from each of 24 complex patterns, some Ss require an excessively and unpredictably long time to complete the test—sometimes more than an hour and a half.

An item analysis was conducted with the aim of reducing the length of time required to administer the test without substantially altering what the test measures. The embedded-figures test was administered to 50 college students, 19 men and 31 women, as a part of a larger study (1). From this group, 26 Ss, 13 from each extreme of the distribution, were selected. Twenty-four single-classification analyses of variance, one for each pattern, were conducted. The 12 items which most significantly (all at better than the .01 level) discriminated between the extreme groups (and hence showed the highest relationship with the whole scale) were selected for inclusion in a short form. The following patterns comprise the short form, using Witkin's (5) numbering system: C-1, D-1, E-1, A-2, C-2, G-1, A-3, H-1, E-3, C-3, D-2, and E-5.

The short form correlates .99 with the whole scale when the original 50 Ss are used. In a cross validation with Witkin's (5) original group,² the short-form—full-scale correlations are as follows: men ($N = 51$), .96; women ($N = 51$), .96. In an additional cross validation with a group of Ss given the embedded-figures test, as part of an as yet unpublished study conducted at The Menninger Foundation, the short-form—full-scale correlations are: men ($N = 30$), .98; women ($N = 30$), .98.

² The author is grateful to Dr. H. A. Witkin for making these data available.

Although reducing the number of items from 24 to 12 effected more than a 40 per cent reduction in administration time, the possibility of a further reduction remained. Since eliminating additional items could reduce the reliability, a reduction in the time limit for each item from five to three minutes was considered. In the Menninger group the correlations between the short form with the three-minute time limit and the original scale are: men, .96; women, .97. The three-minute time limit may have some advantage in that the distribution of items has a less extended tail and is not as asymmetric as the distribution of each item with the five-minute time limit.

It thus appears that the 12-item short form with a three-minute limit gives a very good approximation of results obtained from the entire Witkin embedded-figures test while requiring only half as much administration time.

REFERENCES

1. JACKSON, D. N. Stability in resistance to field forces. Unpublished doctor's dissertation, Purdue Univer., 1955.
2. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren, I: Über den Einfluss gehäufte Einprägung von Figuren auf ihre Sichtbarkeit in umfassenden Konfigurationen. *Psychol. Forsch.*, 1926, 8, 261-317.
3. GUILFORD, J. P., & LACEY, J. I. (Eds.) *Printed classification tests*. Washington: U. S. Government Printing Office, 1947.
4. THURSTONE, L. L. *A factorial study of perception*. Chicago: Univer. of Chicago Press, 1944.
5. WITKIN, H. A. Individual differences in ease of perception of embedded figures. *J. Pers.*, 1950, 19, 1-15.
6. WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. *Personality through perception*. New York: Harper, 1954.

Received September 26, 1955.

ACHIEVEMENT MOTIVATION AND LEARNING

PATRICIA A. KAROLCHUCK

AND

LEONARD WORELL

Portland, Oregon

Reed College

THE purpose of this study was to investigate the effects of n Achievement on two kinds of learning, directed and incidental. A successful attempt was recently made by Lowell (1) to predict degree of learning from the McClelland n Achievement index. He found that a group with high n Achievement increased its output more from beginning to end of a Scrambled Words Task than a group with low n Achievement. Our intent

here was to re-examine Lowell's findings using altered learning materials and to extend our investigation to the effects of n Achievement on incidental learning. From the hypothesis that increased achievement motivation leads to increased learning, we anticipated that a high n Achievement group, as compared with a low n Achievement group, would manifest a greater degree of both directed and incidental learning.

METHOD

Materials. The learning task consisted of an adaptation of a James Thurber short story which was unfamiliar to the subjects (Ss). A multiple-choice test was developed to determine the degree of learning of the story. A similar test was employed by Postman and Senders (3) who investigated the influence of set in a learning situation by providing instructions which centered attention upon either general or specific aspects of learning material. They assumed that the effects of an explicit set were tested when S was asked a question that he had been prepared to expect. Further, when S was questioned on matters unrelated to this set, the effects of incidental learning were being examined. Our multiple-choice test comprised 20 questions, ten each dealing with details of content and general comprehension.

Subjects and procedure. The McClelland group TAT, comprising four pictures,¹ was administered to 108 high school freshmen under neutral conditions. Each picture was projected on a screen for 20 seconds, after which S was allowed four minutes to write a story.

The TAT stories were scored by two judges (inter-rater $r = .96$). The 12 highest and 14 lowest Ss in n Achievement were used to test the hypotheses. These groups did not differ in age, IQ, grade average, or reading ability. Prior to reading the short story, Ss were told to look for spelling and grammatical errors and details of content. In order to insure a single reading of the story, each S recorded his total reading time and handed in his test immediately upon completion.

RESULTS AND DISCUSSION

Contrary to Lowell's findings (1), we have failed to support an anticipated relationship between n Achievement and directed learning. Individuals with High n Achievement did not perform more effectively than Low n Achievement scorers in a situation in which a particular kind of learning was the primary induced goal (Table 1). There was, however, substantial evidence for the presence of a relationship between n Achievement and learning that was irrelevant to the induced set. High n Achievement individuals manifested more learning of an incidental nature than those with Low n Achievement (Table 1).

The discrepancy between Lowell's experiment and the present one may be attributed to a difference in the nature of the learning situations. In

TABLE 1

MEANS AND STANDARD DEVIATIONS OF HIGH AND LOW n ACHIEVEMENT GROUPS IN DIRECTED AND INCIDENTAL LEARNING CONDITIONS

Learning Condition	Low n Ach.		High n Ach.		<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD		
Directed	5.21	9.73	5.75	8.02	.12	.95
Incidental	4.07	9.80	5.83	7.80	1.87	<.05

the Scrambled Words Task utilized by Lowell, performance could be improved by practice and increased effort, while in this study no opportunity for improvement was provided. This explanation, however, does not seem adequate in the light of the occurrence of superior incidental learning by High n Achievement scorers in this research. For the present, we believe that our results must be accepted as empirical findings because of the unclear composition of n Achievement. In re-examining the referents for n Achievement, at least two distinct components appear to be involved, an experience and a motivational variable.² McClelland *et al.* (2) seem to suggest this possibility in referring to motives as comprising expectations, which are "... based on repeated exposure to certain stimuli..." (2, p. 76), and achievement interest or desire. Though the terminology may not be congenial, the essential point remains that several variables may be operant in the n Achievement Scale, whose antecedent conditions are unclear and whose relationship to one another is unspecified. Consequently, it would seem necessary to examine the McClelland index with respect to the nature of the variables involved and to make provisions for their control in order to maximize prediction.

SUMMARY

This study was designed to determine the relationship between n Achievement and learning. It was hypothesized that individuals with high n Achievement would learn more in directed and incidental learning situations than those with low n Achievement.

The findings may be summarized as follows:

² For example, two subcategories, obstacles or blocks (Bp, Bw) and achievement thema (Ach Th), are each given credit toward indicating n Achievement. It would seem to the authors that "obstacles or blocks" may be related to *expectancies* for nonattainment of motive satisfaction, whereas "achievement thema" reflects *concern* over goal achievement but not necessarily expectancy of occurrence.

¹ For a more detailed description of the materials and procedure, see McClelland *et al.* (2). The following four pictures were used:

- two men in a shop working at a machine (3, p. 100)
- boy with an open book in front of him at a desk (3, p. 101)
- an older man and a younger man (TAT 7 BM)
- young boy standing before a surgical mural (TAT 8 BM)

(a) The expected superiority of High n Achievement Ss in a directed learning situation was not found. Our results are not in agreement with Lowell's findings. (b) High n Achievement scorers on the other hand, did demonstrate more efficient incidental learning. The present difficulty in interpreting our results was considered to be a function of the unclear specification of the several possible variables constituting the n Achievement Index, and of their interrelationships.

REFERENCES

1. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31-40.
2. MCCLELLAND, D., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. The achievement motive. New York: Appleton-Century-Crofts, 1953.
3. POSTMAN, L., & SENDERS, VIRGINIA. Incidental learning and generality of set. *J. exp. Psychol.*, 1946, **36**, 153-165.

Received October 11, 1955.

THE CHILD'S PERCEPTION OF THE PARENT

JEROME KAGAN

The Ohio State University

MUCH of the contemporary research on parent-child relationships has centered on the child's perception of his parents. It is generally assumed that acquisition of behaviors like dominance, submission, and other response patterns involved in identification, is influenced by the child's perception of the role-behaviors of each parent figure. However, because of the lack of relevant empirical data on a normal sample it is difficult to (a) assess the significance of the perceptions of a pathological sample and (b) evaluate the relative merit of the various hypothetical statements relating parental behavior to personality development.

Funkenstein, King, and Drolette (1) administered to a group of Harvard students a questionnaire designed to reveal the subject's perception of his parents. The most frequent perception was that of the father as chief source of authority and the mother as major source of affection. It was emphasized that Harvard students are a select sample and caution must be exercised in generalizing these findings. The present brief report summarizes the answers of a much different population, a group of Ohio school children, to direct questions tapping attitudes toward both parents.

METHOD

Interviews were held with 217 children (106 girls and 111 boys), ages 6-0 to 10-2, from grades one, two and three in a Columbus, Ohio, public school. Their answers to the following questions were obtained.

1. If you were in an argument at home with your mother and father, who would be on your side, your mother or your father?
2. Let's make believe you were bad and your mother and father were both home. Who would punish you, your mother or your father?
3. Who is the boss in your house, your mother or your father?

4. Who are you scared of more, your mother or your father?

RESULTS AND DISCUSSION

Table 1 shows the percentage of boys and girls answering "Mother" to each of the four questions. The majority of boys and girls perceived mothers as friendlier, less punitive, less dominant, and less threatening than fathers ($p < .01$ for each question; all p values are for two tails). These results are in essential agreement with those of Funkenstein, King, and Drolette (1).

TABLE 1
PERCENTAGE OF BOYS AND GIRLS RESPONDING
"MOTHER" TO EACH OF THE FOUR QUESTIONS

Group	N	Question 1	Question 2	Question 3	Question 4
Boys	111	60.9	36.4	27.9	16.7
Girls	106	73.6	33.0	19.8	29.2

Although one reasonable explanation of the data is that these answers reflected the child's perception of the objective family situation, it is possible that the statements of both the Harvard students and the Ohio children were influenced, in part, by a culturally reinforced attitude that fathers *ought to be* less affectionate and more dominant than mothers. This hypothesis is tentatively supported by the children's verbal elaborations to the questions. For example, to the question "Who is the boss in your house . . . ?" many Ss hesitated before answering and if they said "Father" often explained their answer by adding "Because he has to work for a living" or "Because he owns the house."

TABLE 2
PERCENTAGE OF BOYS AND GIRLS RESPONDING
"MOTHER" TO THE QUESTIONS IN RELATION TO
GRADE PLACEMENT

Question	Grades 1 and 2		Grade 3	
	Girls	Boys	Girls	Boys
1	75.7	54.9	69.4	66.1
2	28.6	46.3	41.6	27.3
3	12.9	30.4	33.3	25.5
4	27.1	21.9	33.3	12.9

With children of different ages (or grades) combined, differences between the sexes for the four questions were not large. Dividing the group into first- and second-grade children versus third-graders, however, revealed significant differences. Table 2 shows the percentage of the children answering "Mother" to the four questions in relation to their grade-placement. For all four questions the older children were consistently more likely than the younger ones to see the same-sex parent as less benevolent and more frustrating. On Question 2, the third-grade boys viewed the mother as less punitive than the younger boys ($p = .07$). Similarly, on Question 3, significantly more third-grade girls saw the mother as boss than first- and second-grade girls ($p < .05$). One might assume that the parent who is perceived as "boss" is the major controller of gratifications and punishments and, therefore, the one who would be feared more. On Question 4, there was no large difference between the sexes for the younger children, but fear of the mother was significantly more frequent for the older girls than for the older boys ($p < .05$), supporting the trend toward perception of the same-sex parent as more threatening as the child grows older.

This trend toward a more threatening perception of the same-sex parent may have realistic bases and result, in part, from differential handling of boys and girls as they start school and begin to assume more definitive sex roles. In time the child may begin to perceive such a difference in parental attitude and treatment. A recent analysis (2) of interview data from mothers of preschool children showed that the mothers of girls were less nurturant at bed time and more frustrating at meals than mothers of boys of the same age, suggesting that a parent is less gratifying with a child of the same sex than with one of the opposite sex.

SUMMARY

Interviews were held with 217 children of both sexes whose ages ranged from 6-0 to 10-2. Their answers to questions related to their perception of their parents were obtained. Both girls and boys stated that fathers were less friendly and more dominant, punitive, and threatening than mothers. There was, however, a consistent tendency for the older children to be more likely than the younger children to view the parent of the same sex as more dominant and punitive. It was suggested that differential handling of boys and girls might partially account for this latter finding.

REFERENCES

1. FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. E. Perception of parents and social attitudes. Paper read at meeting of Amer. Psychopath. Soc., June, 1955.
2. SEARS, R. R., WHITING, J. W. M., NOWLIS, V., & SEARS, PAULINE S. Some child rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, No. 47.

Received September 13, 1955.

A FURTHER STUDY OF THE MCKINNEY REPORTING TEST SCORES¹

JOHN R. BARRY

Western Psychiatric Institute, University of Pittsburgh

RECENTLY, Douglass and Brown (4) reported on their investigation of reactions to frustration on the McKinney Reporting Test and concluded that the test measures of frustration reactions should be refined further. A study of these

measures, utilizing a similar research design, was consequently undertaken as part of a large personality screening research program (7). The methodology and findings are described below and contrasted with those of Douglass and Brown.

¹ This research was completed at the Department of Clinical Psychology, USAF School of Aviation Medicine, Randolph Field, Texas. Appreciation for their help is due to Drs. S. B. Sells and D. K. Trites, USAF School of Aviation Medicine, and to Dr. Roy M. Hamlin, Western Psychiatric Institute.

The McKinney Reporting Test (5, 6) is a paper-and-pencil group test involving simple repetitive perceptual-motor responses to 315 items. The test consists of four comparable parts: Part I (the pre-frustration condition), Parts II and III (the frustration conditions), and Part IV (the post-

TABLE 1
SIGNIFICANCE OF DIFFERENCES (p -VALUES*) BETWEEN
CORRELATIONS OF PART SCORES IN PRESENT
STUDY AND IN THE DOUGLASS-BROWN
(D-B) STUDY

Part of test	Items attempted		Items correct		Percentage of items correct	
	Present study	D-B study	Present study	D-B study	Present study	D-B study
II	.07*	—	—	—	—	—
III	.03	—	—	—	.001*	.05
IV	.01	—	.01	—	—	—
III	.01	—	—	—	—	—
IV	.05	.05	.001*	—	—	—
IV	.001*	.01	.08	.01	—	—

*All p -values in this table are based upon a two-tailed test of significance. The asterisks indicate that the difference is inconsistent with the hypothesis that stress-sensitive subjects respond consistently from part to part than stress-resistant subjects.

ration condition). Three scores are obtained from each of the four parts: the number of responses made (correct plus incorrect), the number of responses correct, and the percentage correct (i.e., number correct divided by the number made). Douglass and Brown administered the test under the standard (frustrating) conditions to a group of 101 college students, and under nonfrustrating conditions to a similar group. It was hypothesized that the disorganizing effects of the frustrating conditions would be reflected in the lower intraindividual correlations among the part scores of the frustrated group when compared with the analogous correlations in the nonfrustrated group. Fifteen comparisons were made, utilizing the three scores described above. Four comparisons yielded statistically significant differences ($p < .05$) consistent with the research hypothesis. The authors concluded that the evidence for the sensitivity of the test to the frustrating conditions was not conclusive.

PROCEDURE

In the present study, the McKinney Test was administered to 371 first pilots (Aircraft Commanders) and 428 copilots prior to their entering B-29 Combat Crew Training. Criterion ratings of each pilot's adjustment in training were developed to reflect anxiety-proneness, a predisposition to unfavorable reactions to stress and frustration, and generally poor defenses. The ratings were based upon medical and training records, sociometric data, and an eight-hour psychological evaluation, including an interview and psychometric testing. The training records included instructor's grades and comments, board proceedings, and the reasons why a student repeated or failed any parts of the program. These criteria were used by psychologists to make over-all ratings of adjustment on a seven-point scale. Intercorrelations among four raters varied from .85 to .95, indicating a high degree

of rater agreement. These criteria are more completely described elsewhere (1, 2, 3).

A sample of 400 pilots, 200 of whom were upper and 200 lower criterion group members, was selected from the population of 799 pilots in such a way as to define extreme criterion groups. Specifically, the 100 first pilots who were rated as having made the best adjustment to training and the 100 rated as having made the poorest adjustment were designated as members of the upper and lower criterion groups respectively. In the same manner, 200 copilots were selected for the upper and lower groups. Thus, the two criterion groups were selected to define the maximum adjustment differences in the population.

It was predicted that the upper criterion group members would perform on the McKinney Test at a more consistent rate and would be less affected by the frustrating conditions of the test than the lower criterion group. The lower criterion group members were expected to react in an unsystematic fashion during the stress parts of the test. Some of the lower group were expected to perform more efficiently and others more poorly under stress than the upper group. Thus, correlations among the part scores of the upper group were expected to be higher than those occurring in the lower group. Product-moment correlations between the part scores of upper and lower groups were compared by computing the critical ratios between z transformations of the correlations.

RESULTS

Table 1 lists p values indicating the significance of the differences between correlations of the part scores. In the present study, the differences between the criterion groups were in the predicted direction and statistically significant ($p < .05$) for five of the 18 possible comparisons. For three additional comparisons, very significant differences ($p < .01$) in a direction *opposite* from that predicted were found. For two of these three comparisons, the Douglass-Brown data yielded statistically significant differences in a direction *opposite* from that found in the present study. In contrast to the findings in the present study, none of the statistically significant differences found by Douglass and Brown were inconsistent with their predictions. Thus, there appeared to be greater variability in the data of the present study, probably due in part to the criterion differences between the two groups and in part to the differences in the types of setting in which testing was done. In the present study, a partially successful effort was made to create the contrary-to-fact impression that each subject's test responses would have an influence on his subsequent military career. Thus, the test-taking attitudes of the military subjects

*No estimates of stress-sensitivity (unfavorable reaction to frustration) which were independent of the adjustment criterion were available. However, an inspection of the bases of the criterion ratings (3, pp. 4-8) suggested that real differences in frustration reactions should be found between the criterion groups.

probably were different from those of the college students in the Douglass-Brown study.

While it seems likely that the inconsistencies between the two studies may be partly due to differences in the research designs, it is also quite possible that the measures of frustration reaction reflected by both sets of data are unstable. Although certain aspects of the test appear to have frustrating effects,³ various unidentified factors may be masking or interfering with the predicted relationships. The need for refinement in scoring and construction of the test, suggested by Douglass and Brown, is emphasized by these inconsistencies. Since in both studies there are significant differences in the predicted direction between the average intercorrelations of the experimental and control groups, both studies seem to support the experimental hypothesis that the test does enable real discrimination by means of frustration reactions between stress-sensitive and stress-resistant groups. The stress-sensitive persons as a group were less consistent in their test performance than the stress-resistant group.

SUMMARY

It was expected the stress-sensitive group would respond less consistently from part to part on this test than the stress-resistant group. The degree to which this was found is indicated in Table 1. Inconsistencies between the present study and an earlier

³ In research reported elsewhere (1, 2), upper and lower criterion groups, similar to those described above, could be identified accurately by their responses to this test. Specifically, the upper group performed more efficiently than the lower group on all four parts of the test. The examiner-induced frustration in Parts II and III did not appear to sharpen the discrimination between the groups.

one probably reflect the instability of the measure of frustration reaction as well as differences in research designs. There appears to be a need for refinement in the scoring and construction of the test.

REFERENCES

1. BARRY, J. R., FULKERSON, S. C., & SELLS, S. Adaptability screening of flying personnel: Research on the McKinney Reporting Test. *USAF Sch. Aviat. Med. Proj. Rep.* (Rep. No. 56-5) (in press).
2. BARRY, J. R., FULKERSON, S. C., & SELLS, S. Prediction of adjustment with an "object group frustration" test. *Amer. Psychologist*, 1955, 10, 448-449. (Abstract)
3. BITTERMAN, M. E., HOLTZMAN, W. H., & BARRY, J. R. Psychiatric screening of flying personnel: Conditioning and extinction of galvanic skin response as a function of adjustment to flight crew training. *USAF Sch. Aviat. Med. Proj. Rep.* Feb., 1955 (Rep. No. 55-29).
4. DOUGLASS, W. B., & BROWN, K. B. Adequacy of Reporting Test measures for detecting differences to frustration. *J. abnorm. soc. Psychol.*, 1954, 49, 593-595.
5. FULKERSON, S. C. Individual differences in reaction to failure-induced stress. *Amer. Psychologist*, 1955, 10, 336-337 (Abstract)
6. MCKINNEY, F., STROTHER, G. B., HINES, RUTH K., & ALLEE, RUTH A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, 46, 316-323.
7. SELLS, S. B., & BARRY, J. R. A research program to develop psychiatric selection of flying personnel. I. Theoretical approach and research design. I. Research progress. *J. aviat. Med.*, 1953, 24, 29-47.

Received October 13, 1955.

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOEN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHLET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study
in the Behavioral Sciences

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

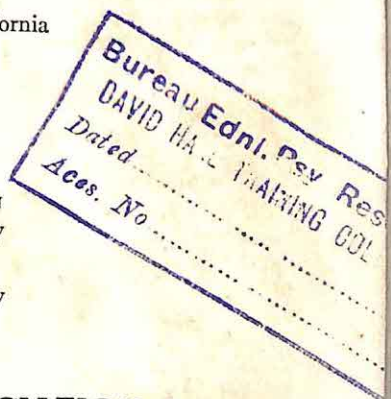
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D. C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 3440 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1956 by the American Psychological Association, Inc.



designate the traditional patterns. The model of the custodial orientations is the traditional prison and the "chronic" mental hospital which provide a highly controlled setting concerned mainly with the detention and safe-keeping of its inmates. Patients are conceived of in stereotyped terms as categorically different from "normal" people, as totally irrational, insensitive to others, unpredictable, and dangerous. Mental illness is attributed primarily to poor heredity, organic lesion, and the like. In consequence, the staff cannot expect to understand the patients, to engage in meaningful relationships with them, nor in most cases to do them much good. Custodialism is saturated with pessimism, impersonalness, and watchful mistrust. The custodial conception of the hospital is autocratic, involving as it does a rigid status hierarchy, a unilateral downward flow of power, and minimal communication within and across status lines.

The newer orientations will be termed "humanistic" (after Fromm (6)) in view of their concern with the individuality and the human needs of both patients and personnel. These orientations conceive of the hospital as a therapeutic community rather than a custodial institution. They emphasize interpersonal and intrapsychic sources of mental illness, often to the neglect of possible hereditary and somatic sources. They view patients in more psychological and less moralistic terms. They are optimistic, sometimes to an unrealistic degree, about the possibilities of patient recovery in a maximally therapeutic environment. They attempt in varying degrees to democratize the hospital, to maximize the therapeutic functions of nonmedical personnel, to increase patient self-determination individually and collectively, and to open up communication wherever possible. While the humanistic orientations have the above characteristics in common, and even more an opposition to custodialism, they still differ among themselves in important respects. For example, the concrete manifestations of humanism will differ, although the guiding spirit may be the same, in a large, architecturally horrendous, financially limited state hospital, as contrasted with a small, well-subsidized, private hospital that accepts only patients regarded as good therapeutic risks.

This inquiry makes no assumptions about the actual therapeutic effectiveness of the various approaches. Our primary concern here is with the nature and the determinants of custodialism and humanism.

THE CUSTODIAL MENTAL ILLNESS IDEOLOGY SCALE (CMI)

The initial, field exploration led to the formulation of the "custodialism-humanism" continuum, the polar extremes of which have been described above. The next step was to construct the Custodial Mental Illness Ideology Scale (CMI), an admittedly crude instrument that had two chief functions in the research: (a) To test the hypothesis that a set of seemingly disparate ideas do in fact "go together" to form a relatively coherent orientation in the individual. A derivative function is to determine whether viewpoints approximating our posited prototypes exist with some frequency within various hospital settings. (b) To provide a quantitative and at the same time meaningful measure that facilitates additional study of the nature of these ideas and their relation to other aspects of the individual and his milieu.

The CMI scale consists of 20 statements, broadly diversified to cover numerous facets of the ideological domain: the nature and causes of mental illness, conditions in the hospital, patient-staff relations, and the like. The items were derived from interviews, conversations, and observations of conferences and everyday hospital life. The scale is presented in Table 1. A more extensive description of the field work and derivation of the CMI scale can be found in Gilbert (7).

Scoring procedure. The subjects were instructed to indicate the degree of their agreement or disagreement with each item on a scale ranging from +3 (strong agreement) to -3 (strong disagreement). The responses were converted into scores by means of an a priori, 7-point scoring scheme. It was intended that a high score represent strong adherence to "custodial" ideology as here conceived, and that a low score represent opposition to this viewpoint. Of the 20 scale items, 17 were regarded as custodial, 3 as humanistic. For the "custodial" items, seven points were given for the +3 response, one point for -3. For the "humanistic" items the

TABLE 1
THE CUSTODIAL MENTAL ILLNESS IDEOLOGY (CMI) SCALE

Item	Mean	DP
1. Only persons with considerable psychiatric training should be allowed to form close relationships with patients.	3.5	2.4
3. It is best to prevent the more disturbed patients from mixing with those who are less sick.	5.0	1.8
5. As soon as a person shows signs of mental disturbance he should be hospitalized.	3.3	4.2
*7. Mental illness is an illness like any other.	2.7	2.7
9. Close association with mentally ill people is liable to make even a normal person break down.	2.0	1.4
11. We can make some improvements, but by and large the conditions of mental hospital wards are about as good as they can be considering the type of disturbed patient living there.	3.0	3.6
15. We should be sympathetic with mental patients, but we cannot expect to understand their odd behavior.	3.2	3.8
17. One of the main causes in mental illness is lack of moral strength.	2.8	3.2
*18. When a patient is discharged from a hospital, he can be expected to carry out his responsibilities as a citizen.	3.0	.5
19. Abnormal people are ruled by their emotions; normal people by their reason.	3.8	4.4
21. A mental patient is in no position to make decisions about even everyday living problems.	3.0	3.1
*23. Patients are often kept in the hospital long after they are well enough to get along in the community.	4.2	-.2
25. There is something about mentally ill people that makes it easy to tell them from normal people.	3.0	2.9
27. Few, if any, patients are capable of real friendliness.	2.2	1.7
31. There is hardly a mental patient who isn't liable to attack you unless you take extreme precautions.	2.5	3.0
33. Patients who fail to recover have only themselves to blame; in most cases they have just not tried hard enough.	1.8	1.5
37. "Once a schizophrenic, always a schizophrenic."	2.3	1.3
38. Patients need the same kind of control and discipline as an untrained child.	3.3	2.4
39. With few exceptions most patients haven't the ability to tell right from wrong.	2.4	2.3
40. In experimenting with new methods of ward treatment, hospitals must consider, first and foremost, the safety of patients and personnel.	5.3	2.3

* Items expressing a "humanistic" position; all others are "custodial."

Note.—The item means and *DPs* are those obtained by a sample of 196 mental hospital personnel in Hospitals C, T, and H. Similar *DPs* have been obtained in other samples of personnel, patients, and visitors. Items are numbered as they appear in the questionnaire, which contained other scales and questions.

scoring was reversed. For convenience in comparing scores from scales differing in length, we shall use the mean per item, multiplied by 10. The possible range is thus 10-70 points.

The CMI scale was initially developed on a sample of 335 staff members (aides, student nurses, nurses, and psychiatrists) in three Massachusetts mental hospitals: Hospital C, a large (1,800 bed) institution dealing largely with "chronic" patients; Hospital T, a Veterans Administration hospital of about the same size; and Hospital H, a small (120 bed) state institution providing short-term active treatment. The range for this sample was 15-52, the mean being 31.3 and the *SD*, 9.5. Comparative data for various subgroup-

ings are presented below (Table 2). The reliability (split-half correlation, corrected by Spearman-Brown formula) was .85, and test-retest correlations on several small groups were of similar magnitude. Table 1 presents the means and discriminatory powers (*DP*) of the individual items.² The *DPs* of all items except numbers 18 and 23 reach the .05 level of statistical significance, and most of them are beyond the .01 level.

The above data indicate that the initial form of the CMI scale has adequate reliability and internal consistency, and they provide a

² *DP* of an item reflects its ability to differentiate between extremely high scorers and extreme low scorers (the upper and lower 25 per cent) on the total scale. It is computed as the difference between the means of the high-scoring and the low-scoring groups.

TABLE 2
CMI MEAN, F MEAN, AND INDEX OF STATUS-CUSTODIALISM

Hospital-Status Unit	N	Index of Status- Custo- dialism	CMI Scale			F Scale			rCMI _F
			Mean	Rank	SD	Mean	Rank	SD	
Attendants at Hospital:									
C	29	12	38.0	12	10.2	46.1	12	13.0	.91
T	51	11	37.3	11	10.3	41.4	11	11.7	.82
H	48	10	32.1	8	9.7	32.5	9	13.0	.77
Student Nurses at Hospital:									
C	66	9	33.4	10	7.7	29.0	8	10.9	.44
T	38	8	33.3	9	6.8	28.9	7	8.9	.25
H	16	7	31.3	6	5.7	28.8	6	9.2	.59
Nurses at Hospital:									
C	14	6	31.3	7	9.9	37.3	10	14.5	.90
T	18	5	22.4	2	5.2	26.7	5	8.7	.53
H	21	4	26.9	5	7.7	25.8	3	14.5	.73
Doctors at Hospital:									
C	6	3	25.8	4	7.0	18.1	1	6.2	.75
T	4	2	21.6	1	4.8	26.6	4	12.8	.80
H	24	1	22.7	2	4.5	19.1	2	8.5	.50
Total Status:									
Attendants	128	4	35.3	4	10.4	39.1	4	13.7	.82
Student Nurses	120	3	33.1	3	7.3	28.9	3	10.2	.41
Nurses	53	2	26.5	2	8.3	27.9	2	12.8	.76
Doctors	34	1	23.1	1	5.3	19.7	1	9.1	.46
Total Hospital:									
C	115	3	33.7	3	8.9	33.8	2	14.5	.67
T	111	2	32.9	2	10.2	34.2	3	12.4	.69
H	109	1	29.0	1	8.7	27.7	1	13.1	.76
Total Sample	335		31.3		9.5	31.9		13.7	.71

basis for further improvement. They suggest, moreover, that a person's stand on any single issue represented in the scale is part of a broader, fairly coherent (though seldom fully integrated) ideology that embraces numerous issues of hospital life.

Data were obtained on two "validation groups" to determine whether the CMI score adequately gauges an enduring ideological conviction. One group, containing 10 administrators at Hospital H who are known for their advocacy of humanistic policies, earned a CMI mean of 18.8, with an *SD* of 6.1. The second group comprised the professional staff at the Social Rehabilitation Unit, Belmont Hospital, England, (9) and would also be expected to have a low CMI mean. The obtained mean was 22.9, the *SD*, 6.7. These findings offer additional evidence of scale validity.³

³ In a study to be published shortly, we have obtained correlations of .5 and .8 between the CMI scale and a measure of custodialism in role performance, in two samples of hospital aides. Since at least a moderate

PSYCHOLOGICAL BASES OF CUSTODIALISM AND HUMANISM

With the CMI scale developed, it was possible to test the hypothesis that *the custodial orientation is one facet of an authoritarian personality, the humanistic orientation a facet of an equalitarian personality*. Several lines of theory and observation led to this expectation. Custodialism is strongly autocratic in its conception of the hospital and ethnocentric in its conception of patients as an inferior and threatening outgroup entitled to few if any of the rights of "normal" people. Humanism, on the other hand, seeks a more democratic hospital structure and regards patients as individuals to be understood and treated rather than as an outgroup to be pitied or condemned. There is considerable evidence that autocratic viewpoints tend to exist within authoritarian personality structures (1, 5, 6).

relation between ideology and action would be expected on theoretical grounds, evidence of such a relation has indirect validation relevance for the CMI scale (4).

We accordingly predicted that the CMI scale would correlate significantly with the F scale (1), a relatively nonideological measure of authoritarianism, and with the Traditional Family Ideology (TFI) Scale (11), a measure of autocratic ideology regarding issues such as husband-wife and parent-child relations.⁴

The obtained interscale correlations are as follows: CMI and F correlate .67, .69, and .76 in Hospitals C, T, and H, respectively. The comparable correlations between CMI and TFI are .50, .56, and .77. The respective *N*s in the three hospitals were 115, 111, and 109. The scale means and *SD*s are presented in Table 2. The findings lend support to the hypothesis that an individual's views regarding mental illness and the hospital are imbedded within a broader ideological and psychodynamic matrix.

It may be argued that the F scale is made of the same stuff as CMI, that it taps relatively superficial ideas or values rather than more central aspects of personality. If this be true, then the foregoing inferences concerning the psychodynamic bases of ideology are unjustified. It is certainly possible that a person may accept many of the ideas represented in the F scale without being an "authoritarian personality." However, we propose on both theoretical and empirical grounds that such persons are rather the exception than the rule. The F items taken as a whole do not comprise an organized body of doctrine. The obtained consistency of response to these items is, we believe, determined for the most part by an enduring pattern of intrapersonal dispositions. Empirical support for this view is given by Adorno, *et al.* (1) and others; for a critical summary, see Christie (3). Significant relationships between CMI scores and nonscale measures of authoritarianism have been obtained by Gilbert (7) and Pine (13). These studies utilized content analysis of interviews, TATs, open-ended questions, and the like in assessing authoritarianism. The F scale would seem to provide a relatively valid though by no means infallible estimate of personal authoritarianism.

⁴ An abbreviated F scale of 8 items was used; it contained Items 9, 13, 18, 25, 26, 37, and 42 from Form 45, and Item 32 from Form 78 of the original F scale (1). The TFI measure contained Items 2, 3, 5, 6, 7, 9, 11, 12 of the short form presented by Levinson and Huffman (14, p. 268).

Custodial ideology has important psychic functions for authoritarian hospital members. The idea that patient behavior is simply irrational and not understandable has great value in reducing inner strain and maintaining self-esteem for personnel who have difficulty at the outset in taking an intraceptive, psychological approach. Again, for the person who has a great defensive need to displace and project aggressive wishes concerning authority figures to those who can be regarded as immoral, custodial ideology has special equilibrium-maintaining value through its justification of punitive, suppressive measures.

Humanistic ideology has corresponding functions for its adherents. By supporting a critical attitude toward the established order, it permits many equalitarian individuals to express generalized anti-authority hostilities in an ego-syntonic form. The principle of "self-control through self-understanding," applied in the treatment of patients, often serves to maintain and consolidate the intellectualizing defenses of equalitarian personnel. In our view, then, both custodialism and humanism have important nonrational functions for their proponents.

RELATIONSHIPS AMONG IDEOLOGY, PERSONALITY, AND HOSPITAL POLICY

We have been concerned thus far with ideology as an aspect of personality. We have suggested that the individual's orientation to mental illness is an intrinsic part of his general approach to life problems and is related to deeper-lying personality dynamics. This, however, is only part of the story. Ideology is also an aspect of the social milieu; we must consider both the psychological and the social soil within which ideologies are formed and modified.

Various social factors operate to induce some degree of ideological uniformity among members of a given occupational status, as well as among members of a total hospital system. Many social scientists, including psychologists, argue or implicitly assume that ideological conformity is ordinarily achieved and that some sort of J curve or concentration of viewpoints approximating the institutional requirements will be found among members of a given institution. One serious limitation of this approach, in our view, is its neglect of the

part played by personality. We would expect that the achievement of a policy-congruent modal ideology depends in part on the presence of a corresponding modal personality. Conversely, to the extent that there is variability in ideology-relevant personality characteristics, we would expect ideological variability among members of any system.

Our three domains of inquiry are ideology, personality, and system requirements. Within each domain we have measured individual or system differences along a given continuum: (a) The custodial-humanistic continuum of individual ideology, as measured by the CMI scale, (b) The authoritarian-equalitarian continuum of personality, as measured by the F scale, (c) The third continuum, custodialism-humanism in system requirements, which was assessed as follows. The sample contained 12 subsystems: four occupational statuses (aide, student nurse, nurse, and psychiatrist) in each of three hospitals. Our procedure was to rank the 12 systems in order from relatively most custodial to relatively most humanistic with regard to the demands and pressures each system placed on its members. We ranked the hospitals first, then the statuses, and then combined the two sets into one series of 12 ranks.

The hospitals were ranked in terms of their degree of change away from a predominantly custodial emphasis on protection and bodily care of patients. The large state hospital, C, was assessed as the most custodial in view of its structural emphasis on detention, protection, and custodial care of patients in a highly controlled setting. The pressures it exerted on personnel, and the kinds of experience it offered them, seemed most conducive of a custodial orientation. The large VA hospital, T, was considered intermediate or transitional in that it was in process of fairly rapid change away from custodialism. The third hospital, H, was the most humanistic of the three in its program of ward care, patient government, and general staff-patient relationships.

The four statuses were ranked in degree of custodialism on the basis of educational level and job requirements vis-à-vis the patient. In order from high to low in degree of custodialism, they fall as follows: aide, student nurse, nurse, and psychiatrist.

The three hospitals and the four statuses were then combined into a series of 12 hospital-

status units ranked according to degree of pressure toward custodialism. Since occupational status pressures operate over a longer period of time, and more selectively, than do hospital pressures, we made status a primary basis of stratification, and hospital secondary. That is, we assumed that statuses are relatively nonoverlapping in degree of custodialism in their policy requirements and that hospitals make a difference only within a single status grouping. Accordingly the rank 12 was given to the most custodial status in the most custodial hospital, namely, the aide status at Hospital C; this is followed by the aide status at Hospital T, and at H; then come the student nurse statuses at C, T, and H; the nurse statuses at C, T, and H; and lastly the doctor statuses at Hospitals C, T, and H with ranks of 3, 2, and 1 respectively (see Table 2). Ideally a more intensive sociological analysis of the structure and policies of each status in each hospital should be carried out. However, the rankings used here seem adequate for our present purposes.

Having roughly assessed the degree of custodialism in the policy requirements of each hospital-status system, we can now investigate the degree to which these requirements are supported by the ideologies, and are congruent with the personalities, of the system members.

Relations between policy requirements and ideology. What is the relationship between the degree of custodialism in the policy requirements of a hospital-status system and the degree of custodialism in the modal ideology of its members? The relevant data are given in Table 2. We use the CMI mean as a measure of modal ideology, for in the distribution of CMI scores the mean, by and large, corresponds closely to the mode. The obtained rank-order correlation between degree of custodialism in policy requirements (status ranks) and in modal ideology (CMI mean) is .92. There is, in other words, relatively great congruence between policy demands and modal ideology. At the same time, the CMI means of the 12 status units do not correspond fully in absolute degree to the estimated degree of custodialism in their structural pressures. For example, the aide status at Hospital C was ranked most custodial both in policy requirements and in CMI mean; however, in an absolute sense, the policy requirements are

highly custodial whereas the CMI mean is only moderate.

The above findings do not tell us how much ideological variability exists within each system. Data on variability are given in Table 2. It will be noted that the *SDs* of most of the 12 units approximate the *SD* for the total sample. Only in the doctor statuses is there anything approaching uniformity of opinion. Thus although *modal* ideology is fairly closely related to policy requirements, the findings on intrasystem variability suggest that an individual's ideology does not reflect in a simple way the demands of his occupational milieu.

In investigating the relationship between *individual* ideology and policy requirements, we consider system pressures as characteristics of the individual. Every individual in the sample of 335 was assigned an index figure representing the relative degree of custodialism in the policy requirements of his particular hospital-status unit. This index figure is simply the rank of the individual's status within the series of 12. For instance, each doctor at H, the least custodial status, is assigned an index figure of 1, and each aide at C is assigned an index of 12.

The obtained product-moment correlation⁵ between CMI score and Index of Status-Custodialism is .47. This finding is evidence of a significant but moderate relationship between individual ideology and system pressures. If system pressures were the most weighty determinants of individual ideology, relative ideological homogeneity within statuses should follow, and thus a high correlation (of the order, .7 to .8) between an individual's CMI score and the degree of policy-custodialism of his work unit. However, the degree of uniformity within any system is not as great as a system-centered mode of thinking would require. An individual's ideology can be predicted with only fair accuracy on the basis of his occupational-hospital membership.

Relations between policy requirements and personality. If the individual's ideological orientation is thought to be simply and directly a result of pressures from his work milieu, relatively independent of his personal-

ity, one would not expect the degree of custodialism in system policy to be significantly related to the degree of authoritarianism in modal personality. Rather, the 12 units might be expected to show similar degrees of authoritarianism, as measured by the F-scale means.

In our view, however, some congruence is to be expected between the policy requirements of a system and the modal personality of its members. Such congruence would be facilitated through recruitment, selective turnover, and possible personality changes in the direction of congruence. We are supported in this hypothesis by the finding of congruence (the correlation of .92) between policy requirements and modal ideology in the 12 status units. We would expect a parallel correspondence between policy requirements and modal personality.

For the 12 status units, the obtained rank-order correlation between Index of Status-Custodialism and F mean is .90 (see Table 2). Thus, there is relatively great congruence between policy demands and modal personality. This congruence is as great as that between policy demands and modal ideology.

The obtained correspondence between modal personality (F) and system requirements is accompanied by appreciable variability on the F scale within most of the statuses (Table 2). The size of the variance on F tends to covary with that on CMI ($r = .61$). This leads us to consider the degree to which system membership and personality are related in the individual. We would expect that the correlation found above between status membership and individual CMI score (.47) will hold as well for index of status membership and F score. The findings bear out this prediction. The correlation between the individual's F score and the Index of Custodialism for his status membership is .46.

Relations between ideology and personality. One of our fundamental postulates is that an individual's ideological orientation is intimately bound up with his deeper-lying personality characteristics. We therefore hypothesize, at the collective level, relative congruence between modal ideology and modal personality. The obtained rank-order correlation between CMI mean (our measure of modal ideology) and F mean (our measure of modal personality) for the 12 units is .81. Thus, the congruence between modal personality and

⁵ The use of indices based on rank in a product-moment correlation involves the assumption of equal intervals between ranks. This constitutes a possible source of error, but probably not a great one.

modal ideology in a system is relatively great. As noted earlier, the size of the variance on CMI is also associated with that on F.

With the individual hospital member as the focus of analysis, the CMI-F correlation for the total sample of 335 (regardless of specific status membership) is .71. We can now consider the relationship between ideology and personality when system membership is held constant. The F-CMI correlations for the single status groupings are presented in Table 2. They average .71, a value identical to the CMI-F correlations for the sample as a whole, and 11 of the 12 correlations are significant at the .05 level or better.

We thus have evidence that the differences in modal ideology among the 12 status units are closely related to differences in modal personality. When we find individual differences in ideology within a single status unit, these differences are closely related to differences in personality characteristics.

The theoretical formulations and results presented here concerning the mental hospital have their parallels in other social settings such as the school, the prison, industry, and the family. In all these institutions a small "administrative" elite has the power and responsibility to set goals and to control the destiny of a massive "membership." This larger population, whether children, patients, or prisoners, is a potential threat to society's values; various measures of education and social control are necessary. One of the major forms of conflict arising in these institutions is that between autocratic and democratic orientations. There is considerable evidence from both the present research and related studies that the autocratic-democratic ideological continuum is one aspect of a broader authoritarian-equalitarian personality continuum. Social ideologies have, to a considerable extent, a psychological basis in the personalities of their adherents. A socio-psychological approach provides, we believe, an important adjunct to historic-sociological approaches in the study of ideology.

SUMMARY AND CONCLUSIONS

This inquiry has taken as its starting point the distinction between "custodialism" and "humanism" in the mental hospital. These terms refer to two contrasting ideological orientations and to the corresponding forms

of hospital policy. We have investigated ideology both as an individual and as a collective phenomenon—or, more accurately, we have used both individual and collective modes of analysis in the study of ideology. With regard to the individual, we have tried to assess ideology by means of a specially devised CMI (custodialism-humanism) scale, and to determine the relationships between ideology and other individual characteristics such as psychodynamics and membership in various groups. With regard to the collective unit (e.g., hospital or occupational status), we have tried to assess the degree of custodialism in its policy requirements and in the modal ideology of its personnel, as well as the degree of authoritarianism in the modal personality of its personnel, and to determine the relationships among these.

In the individual, preference for a custodialistic orientation is part of a broader pattern of personal authoritarianism. Correlations averaging about .70 were found between the Custodialism (CMI) scale and the scales measuring autocratic family ideology (TFI) and general authoritarianism (F). Although various hospital groupings differ significantly in mean CMI score, there are appreciable individual differences within most of the groupings studied. These ideological differences within single hospitals and occupations are quite closely related to differences in personality.

In the collective unit, we found relatively great congruence between prevailing policy, modal ideology, and modal personality. The hospital-status units having the most custodial policy requirements had as well the most custodial modal ideologies and the most authoritarian personalities. At the same time, it should be noted that the correspondence among policy, ideology, and personality is far from complete. Each of these aspects of collective life can vary to some extent independently of the others, and the phenomenon of incongruence is as important as that of congruence.

Although none of our groups can be regarded as ideologically homogeneous, some of them showed relatively small dispersion in CMI scores. These groups had a similar dispersion in F scores, and had low CMI and F means. Our data do not tell how the low diversity and the high ideology-personality congruence came about, but they point up

the need for answers to at least the following questions. To what extent do relatively homogeneous systems maintain themselves by recruitment and selective maintenance of individuals whose personalities are receptive to the structurally required ideology? To what extent do systems change the personalities which initially are unreceptive to the prevailing policy? Under what conditions can a system induce most of its members to support the required ideology even when this ideology is personality-incongruent? Under what conditions can the "incongruent" members change the system to a personally more congenial form?

Systems characterized by relatively great ideological diversity were very common in our sample. Moreover, the ideological diversity went hand in hand with diversity in personality, the standard deviations on CMI correlating .61 with those on F. We incline to the belief that significant heterogeneity of opinion and of personality obtains in the majority of institutional settings within modern societies undergoing rapid technological and educational change.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

2. BETTELHEIM, B. *Love is not enough*. Glencoe, Ill.: Free Press, 1950.
3. CHRISTIE, R. Authoritarianism re-examined. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of "The authoritarian personality"*. Glencoe, Ill.: Free Press, 1954.
4. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
5. DICKS, H. V. Personality traits and national socialist ideology. *Hum. Relat.*, 1950, 3, 111-154.
6. FROMM, E. *Man for himself*. New York: Rinehart, 1947.
7. GILBERT, DORIS C. Ideologies concerning mental illness: A sociopsychological study of mental hospital personnel. Unpublished doctor's dissertation, Radcliffe Coll., 1954.
8. GREENBLATT, M., YORK, R., & BROWN, ESTHER L. *From custodial to therapeutic care in mental hospitals*. New York: Russell Sage Foundation, 1955.
9. JONES, M. *The therapeutic community*. New York: Basic Books, 1953.
10. LEIGHTON, A. H. *The governing of men*. Princeton: Princeton Univer. Press, 1944.
11. LEVINSON, D. J., & HUFFMAN, PHYLLIS E. Traditional family ideology and its relation to personality. *J. Pers.*, 1955, 23, 251-273.
12. MAYO, E. *Human problems of industrial civilization*. Cambridge, Mass.: Harvard Univer. Press, 1933.
13. PINE, F. Conceptions of the mentally ill and the self: A study of psychiatric aides. Unpublished doctor's dissertation, Harvard Univer., 1955.
14. STANTON, A., & SCHWARTZ, M. *The mental hospital*. New York: Basic Books, 1954.
15. TANNENBAUM, F. *Crime and the community*. New York: Columbia Univer. Press, 1951.

Received December 1, 1955.

PERSONALITY STRUCTURE AND GROUP STRUCTURE: AN INTERPRETATIVE STUDY OF THEIR RELATIONSHIP THROUGH AN EVENT-STRUCTURE HYPOTHESIS¹

ARNOLD S. TANNENBAUM

Survey Research Center, University of Michigan

AND FLOYD H. ALLPORT

Maxwell Graduate School, Syracuse University

THE RESEARCH to be reported in this article was designed, in part, to test the hypothesis that an individual's pattern of personality constitutes a basis from which one can predict his attitudes toward a work-program in which he is a participant. That there is such a relationship seems almost a matter of common sense. If the group situation is suited to the worker's personality, his attitude toward it, other things being equal, would be expected to be favorable; if it is unsuited, his attitude will probably be unfavorable. For such an hypothesis only confirming results would be anticipated.

Obvious as the above hypothesis may appear, however, the fact that some theoretical guidance is necessary for its testing is soon realized. The notion of "suitedness," for example, mentioned in the opening paragraph, is rather vague. In this problem we have the analysis of the collective work situation, the pattern of a worker's personality, and the degree, or intensity, of each of the characteristics that comprise that pattern. How can we proceed to derive from these items a formula that will show the degree of compatibility or suitedness between the collective situation and the worker's personality? Is suitability a quality, or can it be expressed, physiologically, in terms of energies? If suitedness means that the situation gives the worker "a chance for self-expression," how can self-expression be defined and measured? Then

too, personality is a complex affair; it has many characteristics, each at its own level of intensity. How are these complexities to be accommodated and appraised in predicting the attitude? And once we have operationalized suitability, how shall we conceive the dynamics of attitude to be organismically related to it?

It is to questions such as these that the present article is addressed; and for a conceptual background and quantitative formulation that may be helpful in answering them the writers have made use of the theory of event-structure.

EVENT-STRUCTURE THEORY AND ITS RELATION TO ATTITUDES: LOGIC OF THE EXPERIMENT

A complete summary of the theory of event-structure will not be attempted here; we shall present only a few of the pertinent concepts.² A basic postulate of the system is that behavior can be described as a cycle of events or as a system of such cycles in interrelationship. The contention that the behavioral pattern is a *true cycle* is based on the fact (or likelihood) that the events which make it up return to the region in which they started. For example, "drinking" usually starts with drying in the mouth or throat, and the events that ensue (involving receptors, neurons, muscle fiber contractions, etc.) bring the cycle of ongoings back to the same region, as the membrane of the mouth and throat is moistened. This feature of return to the starting region is known as "closure." The throat-drying represents a "*primary*" event (really many events in a region), which constitutes the initial energetic upset of the cycle. The act of drinking

¹ This study was conducted as part of a larger experiment directed by N. C. Morse and E. Reimer and constituting one of a series of researches, under the Human Relations Program of the Survey Research Center, on organizational functioning. The writers wish to thank Dr. Morse and Mr. Reimer for their generous help in the conduct of this study. They also wish to acknowledge a grant from the Rockefeller Foundation which contributed to its support.

The material in the present article was drawn from a dissertation submitted by A. S. Tannenbaum in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Graduate School of Syracuse University, February, 1954.

² For more complete details of event-structure theory, so far as it is at present formulated, the reader is referred to two publications by F. H. Allport (3, 4). The reader is also referred to a mimeographed document which discusses in greater detail the theory as it pertains to the present research (5).

involves a displacement or equalization of energy through the cycle which reduces the energy of the primary region. It is hypothesized that all behavioral acts, overt or covert, can be described in terms of such self-closed cycles or structures.

Event-structure theory, which is a general conception, provides for the "group" or collective situation in the same way. It is conceived as a structure of a higher order that is built, on the same principles, from the structures of individuals' behaviors. Events occur between individuals; and, depending on probabilities, the events proceed from energetic upset in one part (or individual) around the collective cycle and back to the initial region, in which an intraorganismic equilibrium of the participant is then attained. We have, therefore, a *collective* structure (in our present case, the work group) as well as the individual behavior-structures that compose it.

Another important feature of the theory is its interpretation of meaning. It is hypothesized that a behavior cycle, with repetition, can come to be abridged in energies and in the number of events that involve contacts with the environment. It may be limited to the proprioceptive circuits, or even further restricted to the cerebrum or to the cortex alone. Such a cycle represents the "meaning" of the original overt act. It can also represent such acts in a "generalized" way. It still, however, is a true cycle, having a primary event region and a closure which occurs as the cycle completes itself and comes to equilibrium. Subjectively we say, at such a time, that a "meaning" is experienced.

An aspect of personality is just such a "meaning-cycle." It is one that is characteristic of the individual. We shall call such a personality-characteristic a *trend-cycle* or *trend-structure*. In more familiar language, a trend-cycle refers to something which the individual is "characteristically trying to do" (though not necessarily consciously trying), that is, to some meaning that he is continually seeking to attain or realize through his behavior (1). Its energy level indicates "how hard" he is trying to achieve that meaning. The individual, however, has *more than one* trend; and their cycles are conceived as forming a trend-system. In terms of the theory, it is this system which must be considered as basic in the influence of an individual's personality upon his attitudes

toward the collective situation. One of the variables concerned in the predictive formula of our problem will be the energies which we refer to as the "potency" of the personality trends of the subject.

A trend-cycle usually achieves closure not by itself, but through some (overt) behavioral cycle which affords the "means," that is, through a pattern of events, by which the closure of the trend can be accomplished. Such an "instrumental" cycle or set of cycles may be provided by a collective structure. To illustrate this latter point, let us suppose that a certain player in a volley-ball game has a strong trend-cycle of "trying to cooperate." The collective structure of the game consists of a series of events involving close teamwork on the part of each player; and this fact gives our subject, as he operates in the structure, a continual opportunity for the closure of this trend. The trend cycle and the game-playing cycle are here said to be "interstructurant." We say, more familiarly, that the game affords an opportunity for "expressing this characteristic (or meaning) of his personality."

But there is still another feature to be considered. Not all overt or collective structures give the same opportunity, or means, for the closure of a given trend-cycle. We have therefore to consider the "*degree of interstructurance*" between the trend-cycle and the tangent collective structure.³ We shall here refer to this aspect as the *relevance (R)* of the collective structure to the trend.

Finally we come to the question of attitudes, our dependent variable. An attitude is considered to be a reaction of favor or disfavor toward a social object or symbol, in this case toward the collective work structure. An attitude is energetically bound up with both the individual's trend cycles and the collective structure in which he is involved. Let us suppose that a worker has a strong trend of "trying to rely on others." Now if the organization of the group work-program is such that it requires continual close supervision and direction by a group leader, the program will be highly relevant to this trend; and we might expect that there would be a considerable probability of the trend's closure. This circumstance, combined with the high potency of the trend-cycle, would be expected to add a con-

³ Symbolized as "I" in the references cited.

siderable increment to the favorableness of the worker's attitude toward the program. But let us take the case of another worker who has a trend "to be treated by others as an equal." The types of work-setting just mentioned would probably yield little if any relevance to the trend, and the energetic increment to attitude from this trend would be negligible, or perhaps even negative. Suppose that the nature of the program were such as to leave most decisions to the workers, with little or no supervision. Here the relevances would be reversed for the two individuals, and the predictions for attitude would be correspondingly altered.

The specific illustrations given above compare the attitudes of two different individuals. It must be remembered, however, that the same individual may have trend-cycles (of differing potencies) to which the work situation has different degrees of relevance; and the individual also has many trends. *Summation* of all these energetic effects will therefore be required.

The task of predicting the attitudes of an individual toward a collective structure in which he is operating, then, must provide at least for the following procedures: (a) the obtaining of a set of measurements of his trend-potencies and of the degree of relevance of the collective structure to each of his trends; (b) the combining of these two measurements in some way for each trend; and (c) the combining of all the resulting quantities into some over-all formula by which the individual's attitude can be predicted. We shall be concerned first with the problems involved in the last two steps.

DERIVATION OF A PREDICTIVE FORMULA

As stated earlier, the amount of closure of the trend, a factor upon which the related attitude variable is to be predicted, can be defined as the effect of the collective structure in reducing the primary-event energy of the trend and bringing that structure to equilibrium. Such a measure of the closure effect would be given by subtracting from the trend's initial energy (potency) the amount of energy left in the trend-structure after some standard period of performance of the individual in the collective structure. This quantity can be expressed by

$$C = P_1 - P_2, \quad [1]$$

where C represents the amount of the trend's energy closed; P_1 the amount of the trend's energy at time 1, before closure; and P_2 the amount of its energy remaining after the collective structure's operation (i.e., after closure).

It is our assumption that attitudes can be understood in terms of the elements of this formula. However, we would predict that an attitude toward a collective structure would be a function, not only of the amount of closure afforded by the collective structure, but a negative function of the amount of trend energies remaining unclosed as a result of the operation of the collective structure. The favorableness of the attitude, then, would be predicted to be proportional to the amount of closure minus the amount of unclosure.⁴ Using the symbol A_o to indicate the predicted energy of the attitude toward object o (the collective work situation), and remembering that we are dealing with the increment to the attitude given by only one trend, we have the formula

$$\text{Increment to } A_o = (P_1 - P_2) - P_2. \quad [2]$$

Whenever the value of this equation is negative we might suppose that the direction of the attitude-increment is negative, that is, unfavorable to Object o . But the total energy of the attitude is made up of a summation of the increments from all the trends. Using a subscript T to represent each of the trends in the total personality structure, taken one at a time in the computation and summation, the formula then becomes

$$A_o = \sum (P_{1T} - P_{2T}) - \sum P_{2T} \quad [3]$$

$$= \sum (P_{1T} - 2P_{2T}).$$

And again, where $\sum (P_{1T} - P_{2T})$ (closure) is less than $\sum P_{2T}$ (energy remaining in the trend), thus making the right side of the equation negative, we might predict the attitude to be unfavorable to o .

Equation 3 deals only with the trend-potencies of the individual. We must now return to the task of bringing the relevance of the collective structure to these trends into the quantitative picture. In terms of the theory, both these factors help to determine the favorableness of an attitude. As will be presently shown, the relevance of the collective structure

⁴ For an earlier application of this formula see N. C. Morse (8).

to a trend-structure was experimentally defined in the present study, and estimated as the probability that there are means available in the collective structure for the closure of the trend's energy. That is, if the probability of complete closure were rated as 1.0, it would mean that the rater felt certain that this energy would be fully closed—100 per cent of the trend's energy would be reduced by the collective structure. If the probability of complete closure were rated as 0, one would expect that even though some of its energy might be closed, the proportion would be minimal. Other probability ratings, between 0 and 1.0, could be taken as representing the likelihood of proportions of closure that lie between the minimal and the complete. If, then, R represents the *proportion* of the trend's energy (P_1) which becomes closed, the *amount* so closed becomes P_1R : the postulated relation is multiplicative. Using C for amount of closure, t for a particular trend-structure, and W for the collective work structure, we have the following formulation:

$$C = P_{1t}R_{W \rightarrow t}, \quad [4]$$

or, in terms of the many trends which make up the total personality structure,

$$C = \sum P_{1T}R_{W \rightarrow T}, \quad [5]$$

where C is total amount of closure and T , as before, is a class symbol for the trends whose respective PR s enter into the summation.

It will be remembered that the relation hypothesized between trend-energies, alone, and attitudes was

$$A_o = \sum (P_{1T} - P_{2T}) - \sum P_{2T}, \quad [3]$$

and that closure, on a purely trend-energy basis, was defined as

$$C = (P_1 - P_2). \quad [1]$$

For a number of trends the total amount of closure, on this basis, would be expressed by

$$C = \sum (P_{1T} - P_{2T}). \quad [6]$$

From 5 and 6 we have the equality

$$\sum (P_{1T} - P_{2T}) = \sum P_{1T}R_{W \rightarrow T}. \quad [7]$$

Substituting from 7 into 3

$$A_o = \sum P_{1T}R_{W \rightarrow T} - \sum P_{2T}. \quad [8]$$

But since

$$\sum P_{2T} = \sum P_{1T} - \sum (P_{1T} - P_{2T}) \text{ (identity),}$$

or

$$\sum P_{2T} = \sum P_{1T} - \sum P_{1T}R_{W \rightarrow T},$$

Equation 8 now becomes

$$A_o = 2\sum P_{1T}R_{W \rightarrow T} - \sum P_{1T}. \quad [9]$$

The right-hand side of this equation is the independent variable by which the attitude of the worker is to be predicted. The equation now states the structural hypothesis in terms of variables for which suitable measures can be found, viz., the potencies or energies of the subjects' trend-cycles, ratings by judges of the relevance of the collective work-structure to the trend-structures, and the attitude-scale positions of the subjects. The operations for measuring these variables will be discussed in the following section.⁵

METHOD

Work-Structures and Subjects

Two contrasting types of collective structures were available for the testing of this hypothesis. These were group situations approximating respectively the democratic and autocratic forms of organization. The investigation was conducted in one department of a large business organization, and the subjects (S s) were women who were employed in clerical work (7, 9). Each of the two experimental programs established within the department included about 200 persons. The programs were made to differ in the way in which authority was distributed within them. In the working groups of one of the programs, which was called the "autonomous" program, the employees (clerks) were given greater control than they had previously exercised over some of the work conditions. The responsibilities of each employee were also given greater emphasis. Authority previously held by the company officers was delegated to division managers who, in turn, delegated authority to the section heads and employees as a group. The clerks in this program participated in group meetings at which they discussed topics of their own choice. They made decisions regarding how the work would be distributed, length of the recess period, their right to leave the department during working hours, and similar matters. In the second experimental division, which was called the "hierarchical" program, the activities of the clerks were regulated more rigidly. Some scientific management principles were incorporated. Lectures by com-

⁵ In references 3 and 4 of the bibliography another formula with different notation is proposed. It is based upon a more systematic assembling of structures and a slightly different concept of interstructuranc (relevance).

pany experts, work standards, and incentives were included. Decisions were made primarily by upper-level company officers. These two programs, at the employee level, continued for about one year, a time span that will be referred to as the experimental period.

Trend-Structures and the Measurement of Their Potencies (P_{1i})

A standard sample of 26 trends was used for all Ss. Several judges (*vide infra*) considered a range of "possible" trends that were suggested by a systematic canvass of the likelihood that the work-program (autonomous or hierarchical) would have some appreciable relevance to them. Since the programs were opposite in character, trends having high relevance with respect to one of them would usually have low relevance with respect to the other. The outcome for the trends thus selected (Table 1) will be seen to provide an equal number of trends having an $R > .5$ and an $R < .5$ with respect to each of the work-programs. It is not maintained that this method of selection, and the trend-list that resulted, assured a truly representative sample of trends for the individual Ss, but only that certain gross errors of sampling were eliminated by giving high and low relevance an equal chance to play a part in the predictive score.

In the relevance-rating procedure, three judges were employed who were both psychologically trained and familiar with the work situations.⁶ The method included the drawing up of lists of roles which the Ss were considered as playing in their respective programs. Starting with general categories, a level of specificity was finally contemplated that dealt with such work-structure occurrences as "clerk receives help from supervisor," "clerk participates in a group meeting," "clerk reacts socially with another clerk" and so on. In deducing the trends that might be related to these event-patterns, the judges tried to answer the following question: "What is the nature of the trend meaning-cycle which would obtain closure through this specific series of events?" Additional criteria employed were (a) likelihood of the trend's occurrence within the type of population studied, (b) the applicability of the trend deduced to the early years of life (since it is believed that characteristics of personality are as a rule, long standing), and (c) freedom from ambiguity. Twenty-six trends were finally selected for measurement. They are presented in the left-hand side of Table 1.

A questionnaire was designed to measure the strengths of these trends in each of the Ss. Five questions were employed for each trend, examples of which are as follows: For the trend "trying to be independent of control figures" (Item 7 in Table 1), one of the questions was: "How important is it for you to feel that you can run your life without depending upon people who are older and more experienced than you?" For the trend "trying to understand others' viewpoints and feelings," the following "agree" or "disagree" item was one of the five employed: "I don't care too much what other people think and how they feel about things. What they think is not my concern." In one of the five questions for the trend "trying to keep one's role at any given time distinct from other roles," the sub-

jects were asked: "If you are discussing something serious with a friend, how much would you object to her bringing in some humorous comments?" For the trend "trying to adhere rigidly to specific rules or directions," one question asked "How much do you dislike it if directions on a recipe or sewing pattern are only partially laid out?"

Although some variety was introduced in the form of the questions, the majority were to be answered by checking on a 5-step scale that contained appropriate words indicating degrees of frequency, amount, etc., in which the S had the feeling or tendency in question. Care was taken to make the questions general in form and distinct from the company situation; and the questionnaire was administered at a time considerably earlier than the measuring of the dependent, attitudinal, variables. A potency score (P_{1i}) was computed for each trend for each respondent by taking the mean of her scores on the five questions designed to measure that trend. Test-retest reliabilities for the trend-scores in the population (over a one-year period) varied from .38 to .78, with an average correlation for all trends of .55.

The Measurement of Relevance ($R_{W \rightarrow i}$)

Ratings of the relevance of work-structure to trend-structure were made for each trend, and with respect to each of the two work-programs (autonomous and hierarchical). They were made by two of the judges who had participated in the selection of the trends to be rated. In this procedure the relevance-judge estimated the probability that complete closure for the trend would be given during the course of a day's work in the program concerned. The estimates were made on a numerical scale of 1 to 5, but were later translated in terms of a probability continuum ranging from zero to 1.0, zero meaning no probability of complete closure and 1.0 meaning certainty of complete closure. The mean of the judges' ratings for a given trend, with respect to a given program, was taken as the relevance-score of the program to that trend. These relevance-scores, expressed as probabilities, are shown in the right-hand side of Table 1. The judges made their ratings separately, but after having been associated on the study for some time during which some common understandings no doubt developed. The product-moment correlations between the judges' ratings were .83 for the relevance of the autonomous program, and .87 for the hierarchical. The relevance-ratings were, of course, given only upon the basis of a knowledge of the work-program in relation to the trend under consideration, and never with any particular S in mind.

Assigning "Suitability" Scores

An "attitude-prediction" score was computed for each individual by substituting the appropriate values in the right-hand side of the formula, $2\Sigma P_{1i}R_{W \rightarrow i} - \Sigma P_{1i}$ (Equation 9). For the Ss in the autonomous program the relevance ratings of that program to the trends ($R_{1 \rightarrow i}$, in Table 1) were used, and in the hierarchical program, the ratings $R_{11 \rightarrow i}$ were used. The prediction scores, computed in accordance with the formula, may be considered to indicate the "suitability" of the individual's personality structure to her work-

⁶ N. C. Morse, E. Reimer, and A. S. Tannenbaum.

TABLE 1
TRENDS USED IN THE EXPERIMENT, WITH RELEVANCES OF THE TWO WORK-STRUCTURES

Trend The subject is <i>characteristically trying to</i> :	R of Autono- mous Program to Trend ($R_I \rightarrow t$)	R of Hierar- chical Program to Trend ($R_{II} \rightarrow t$)
1. assume responsibility (anticipating, preparing for, and accepting the consequences of one's activities)	.84	.32
2. be creative (develop new patterns; to use one's imagination; to think up new and original ways of doing things)	.72	.00
3. increase the variety (number of types) of satisfactions in any given situation (to enjoy as many different things as possible in the same situation, e.g., to "mix business with pleasure")	.67	.19
4. understand and explain the reasons for one's behavior and opinions (to understand and explain one's feelings)	.82	.32
5. help others	.69	.32
6. understand others' viewpoints and feelings (to know why people feel or act as they do; to understand the reasons for others' behavior)	.82	.32
7. be independent of control figures (to be free of directions and control of persons in authority)	.82	.13
8. express one's considered judgments (to think through things carefully and state one's opinions)	.82	.19
9. base one's actions on one's own critical judgments and evaluations (to carefully evaluate and decide things for oneself)	.57	.07
10. act on the same level as others (to treat and be treated by others as an equal)	.88	.34
11. understand relations between and reasons for things (to understand the basis for things; to figure things out; to know "why")	.75	.13
12. do one's own thinking (not to be suggestable; to decide things for oneself; to draw one's own conclusions)	.75	.38
13. take the initiative with others (to start things; to organize things)	.94	.13
14. avoid emotional involvements (to avoid the expression and reception of emotional acceptance; to keep emotionally distant from others; to avoid being "warm" in relationships with others; to avoid affection)	.07	.63
15. conform to the wishes of control figures (to please persons in authority)	.13	1.00
16. adhere rigidly to specific rules or directions (to follow rules to the letter; to be given directions for doing things)	.32	1.00
17. act inferior to control figures (to be humble in relation to persons in authority)	.13	.64
18. avoid committing oneself (to keep one's opinions to oneself; to keep from getting oneself "out on a limb"; to keep people from knowing where one stands on controversial questions)	.25	.94
19. keep one's roles (behavior categories, relationships) at any given time distinct from other roles (to keep various parts of one's life separated)	.25	.88
20. admire or respect control figures or symbols (to look up to persons in, or symbols of, authority)	.19	1.00
21. be superior to others (to be better than others; to be more outstanding than others)	.25	.98
22. be efficient (not to waste time or effort)	.32	.88
23. be submissive to control figures (to submit to dominant leaders; to obey strong and forceful authority figures)	.13	.63
24. show self-discipline (to show self control; to keep one's behavior disciplined)	.50	.88
25. depend on others (to receive direction from others)	.38	.82
26. obey rules and follow directions	.25	1.00

program structure (I or II) as represented by W in the formula. Those Ss whose suitability scores were above the median of that variable's distribution in their program were called "suited," and were predicted to fall into the upper half of the attitude distribution (i.e., to have relatively favorable attitudes toward their program). Those who were below the median of the suitability distribution in their program were called "unsuited," and were predicted to fall into the lower half of the attitude distribution (i.e., to have relatively unfavorable attitudes). It should be noted that the attitudes here referred to were those which express favor or disfavor toward the work-program in which

the individual is placed. The extent to which the Ss' attitudes toward their program, as independently measured, correspond to this prediction, constitutes the operational test of the hypothesis. More specifically stated, we are testing the hypothesis that a greater proportion of those relatively "suited" to the program in which they are placed will be more favorable (or less unfavorable) to their program than will be the case for those who are "unsuited" to their program.⁷

⁷ The hypothesis is stated in *relative* terms since other factors also would be expected to influence the absolute level of the subject's attitude.

TABLE 2

DESIRED LENGTH OF PROGRAM IN RELATION TO
PERSONALITY SUITABILITY
(Values represent percentage of Ss in each group)

Program	Personality	Relatively Long	Relatively Short	N
Autonomous	Unsuited	36.7	63.3	100% (N = 30)
	Suited	76.0	24.0	100% (N = 50)
Hierarchical	Unsuited	23.1	76.9	100% (N = 13)
	Suited	33.3	66.7	100% (N = 15)
Combined Data	Unsuited	26.2	73.8	100% (N = 43)
	Suited	60.0	40.0	100% (N = 65)

The Dependent Variable: Attitude Structures and Their Measurement

The attitudinal measures were obtained by means of a paper-and-pencil questionnaire administered at the end of the experimental period, about one year subsequent to the securing of the personality measures. Some of the attitude-measures were relatively direct, in that the respondent was asked specifically about her liking for the program. Others were less direct, but still bore on the question of favor or disfavor toward the program. The specific items, together with the method of scoring, will be described in connection with the results. The attitude data shown below were derived from four questions, each of which was designed to measure an aspect of the central dependent variable. The results on these questions are, of course, not independent; they represent simply a number of ways of appraising the attitude-dimension with which we are concerned.

RESULTS

The results are given in the form of 2×2 tables. Three tables are presented for each attitude-item: one for the data from the autonomous program, one for the hierarchical program, and one for both programs combined.⁸ Each of the distributions, for both the independent and the dependent variable, has been dichotomized at a point as close to the median as possible. The terms "suited" and "unsuited," and "relatively" (e.g., "relatively long" or "relatively short") are used in the tables to indicate the two halves of the dis-

tribution. This procedure means that the break in the dependent-variable distribution in one program may be at a point different from that in the other, or different from that in the combined data. Statistical tests for reliability of differences between proportions were computed (6, pp. 75-77).⁹ Since the hypotheses are directional, one-tailed tests were used. Results below the 5 per cent level of confidence are referred to as significant, while results between the 5 per cent and 10 per cent levels are considered "suggestive."

Desired Length of the Program

Each respondent was asked, "How long would you like the program to last?" This was a "write-in" question coded on a 5-point scale from "end right now" (value of 1) to "indefinitely" or "forever" or "always" (value of 5). The results for this item are presented in Table 2.

A large and significant difference is found between suited and unsuited persons in the autonomous program. Seventy-six per cent of the suited individuals wanted the program to last a relatively long time, whereas only 36.7 per cent of the unsuited wanted it to last so long. The z value for this difference is 3.49. In the hierarchical program, although the difference is in the predicted direction, it is not significant.¹⁰ In the combined data, however, it is found that a significantly greater proportion of the suited individuals (60.0 per cent) desire a relatively long extension of the program. The proportion for unsuited individuals is much smaller (26.2 per cent). The value of z is here 3.45.

Satisfaction with the Way the Program Operates

A second attitude-item asked "How satisfied are you with the way things are operating now in your division and section compared to the way they were operating two years ago?"

⁹ The formula employed was

$$z = \frac{p_1 - p_2}{\sqrt{pq \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

⁸ Pooling the data is legitimate, since the two programs represented two different populations who were oriented toward two different attitude objects, and with no overlapping on any of the terms of the predictive equation.

¹⁰ The reasons for the small N 's include the following: the fact that only those subjects were used who had spent a full year in the program and for whom a full series of questionnaires was obtained and matched, the occurrence of failures to answer write-in questions, and irrelevant responses that could not be coded.

TABLE 3
SATISFACTION WITH THE WAY PROGRAM OPERATES
IN RELATION TO PERSONALITY SUITABILITY
(Values represent percentage of Ss in each group)

Program	Personality	Rela- tively Satis- fied	Rela- tively Unsatis- fied	N
Autonomous	Unsuited	31.8	68.2	100% (N = 44)
	Suited	42.2	57.8	100% (N = 57)
Hierarchical	Unsuited	53.9	46.1	100% (N = 26)
	Suited	63.3	36.7	100% (N = 30)
Combined Data	Unsuited	48.6	51.4	100% (N = 70)
	Suited	59.8	40.2	100% (N = 87)

Answers were checked on a 5-point scale ranging from "much less satisfied with the way things operate now," to "much more satisfied with the way things operate now." Table 3 presents the data for this item.

Although the direction of the difference between suited and unsuited is as expected, the result for neither program proves significant or suggestive. When the data for the programs are combined however, a suggestive relationship is found in the predicted direction, with 48.6 per cent of the unsuited and 59.8 per cent of the suited persons being relatively satisfied with the way their program operates now as compared with two years ago ($z = 1.40$).

Degree of Liking for the Program (First Form of Question)

As a further exploration of the Ss' attitudes the following free-answer question was asked: "How would you describe the present program in your division?" The write-in answers were categorized by coders in six groups, arranged on a scale indicating the degree of liking manifested, and ranging from "like it very much," or "it's very satisfactory," or "never better," to "dislike it very much," or "don't like it at all," or "it's very poor." The results are shown in Table 4.

In the autonomous program proportionately more suited individuals (60.0 per cent) appear to have a greater liking for the program than unsuited individuals (37.5 per cent). The z value for this difference is 1.41, a "suggestive" finding. The difference in the hierarchical

program is not significant or suggestive, but when the data are combined we find that a significantly greater proportion (47.2 per cent) of the suited individuals have more liking for their program than was the case for unsuited (23.1 per cent). The z value is 1.94.

Degree of Liking for the Program (Second Form of Question)

An additional question was asked in order to ascertain the degree of liking for the program, this time in terms of the individual's "feeling" about it. The S was asked: "How do you feel about it (the program) now?" This free-answer question was coded for liking manifested on a 5-point scale ranging from a response that

TABLE 4
LIKING FOR PROGRAM IN RELATION TO PERSONALITY
SUITABILITY
(First form of question; values given are percentages)

Program	Personality	Rela- tively Like	Rela- tively Dislike	N
Autonomous	Unsuited	37.5	62.5	100% (N = 16)
	Suited	60.0	40.0	100% (N = 25)
Hierarchical	Unsuited	40.0	60.0	100% (N = 10)
	Suited	36.4	63.6	100% (N = 11)
Combined Data	Unsuited	23.1	76.9	100% (N = 26)
	Suited	47.2	52.8	100% (N = 36)

TABLE 5
LIKING FOR PROGRAM IN RELATION TO PERSONALITY
SUITABILITY
(Second form of question; values given are percentages)

Program	Personality	Rela- tively Favor- able	Rela- tively Unfav- orable	N
Autonomous	Unsuited	22.6	77.4	100% (N = 31)
	Suited	31.8	68.2	100% (N = 44)
Hierarchical	Unsuited	20.0	80.0	100% (N = 15)
	Suited	45.5	54.5	100% (N = 11)
Combined data	Unsuited	65.2	34.8	100% (N = 46)
	Suited	80.0	20.0	100% (N = 55)

indicated "dislike it very much" to a response indicating "like it very much." The data are presented in Table 5.

Neither program yielded significant or suggestive differences between suited and unsuited. When the data are combined, however, a significant relationship in the predicted direction is found. Proportionately more suited individuals (80.0 per cent) feel more favorable toward the program they are in than do unsuited individuals (65.2 per cent). The z value is 1.67.

SUMMARY AND CONCLUSIONS

The attitude of favor or disfavor, developed by workers in a large business organization to two experimental programs with contrasting patterns of allocating authority appears to be a function, as was predicted, of the interaction between the personality structure of the individual and the structure of the work-program in which he is operating. Those individuals, who by their trend-structures (potencies and relevance both being considered) are "suited" to the program they are in, tend to feel a greater satisfaction with it and a greater degree of liking for it than do those whose trend-structures are "unsuited" to their program. Suited individuals also tend to want their program to last longer. Though these differences are not always large and are not typically significant in considering one program by itself, they are (with one exception) in the expected direction. And when the data for the programs are combined, yielding a larger N , the difference is significant for three of the attitude-measures and suggestive for the fourth. It might be noted, also, that these differences occur despite the fact that, because of the homogeneity of the populations, the initial variances in the personality measures were rather small.

In developing the hypothesis tested in this experiment we have made use of the theory of event-structure. It may have occurred to the reader, however, that the results can be interpreted equally well in terms of other current conceptions of behavior, notions which are much simpler than the theory here presented. It might be contended, for example, that the judges of relevance in the experiment were merely successfully appraising the way in which the worker perceives the group process

in relation to the fulfillment of his motivating needs; and attitude in current theory has for some time been regarded as a function of motive-strength and perceived goal-relevance.

The theory, however, does not lose its significance because of the consideration that results in the same general direction could have been predicted by other, more familiar, conceptions. Every theory requires empirical support for its predictions, and two different theories might require, for confirmation, the same direction of results. The advantages of the event-structure approach, we believe, include the following: (a) It offers an explanation of *how* the interaction between personality and group-action might take place. (b) It provides a rationale for mathematical treatment. (c) It is capable of providing generalizations that are broader than attitude or motive, even broader than the field of behavior.

The present test of event-structure theory is limited in scope and other theories might have predicted the same direction of results. In these respects, therefore, the support given it by the data is not unique or conclusive. The results lie in the direction predicted by the theory; but further investigations will be required before the merit of the structural explanation can be properly appraised.

REFERENCES

1. ALLPORT, F. H. Teleonomic description in the study of personality. *Charact. Pers.*, 1937, 5, 202-214.
2. ALLPORT, F. H. An event-system theory of collection action: with illustrations from economic and political phenomena and the production of war. *J. soc. Psychol.*, 1940, 11, 417-445.
3. ALLPORT, F. H. The structuring of events: outline of a general theory with applications to psychology. *Psychol. Rev.*, 1954, 61, 281-303.
4. ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
5. ALLPORT, F. H. *Theoretical statement concerning the relationship between personality structure and group structure*. Maxwell Graduate School, Syracuse Univer., 1956 (mimeographed).
6. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
7. MORSE, NANCY C., & REIMER, E. The experimental change of a major organizational variable. *J. abnorm. soc. Psychol.*, 1956, 52, 120-129.
8. MORSE, NANCY C. *Satisfaction in the white-collar job*. Ann Arbor: University of Michigan Press, 1953.
9. MORSE, NANCY C., REIMER, E., & TANNENBAUM, A. S. Regulation and control in hierarchical organizations. *J. soc. Issues*, 1951, 7, 41-48.

Received November 29, 1955.

STIMULUS GENERALIZATION AS A FUNCTION OF CLINICAL ANXIETY¹

GERALD ROSENBAUM

Wayne University

THE concept of anxiety has been widely employed by psychologists to explain a variety of behavioral phenomena, ranging from the development and maintenance of various symptomatologies to the learning of the conditioned avoidance response. The clinical notion of anxiety generally corresponds to that of an implicit state or process within the organism which is inferred or defined on the basis of the observation of certain manifest symptoms. On the other hand, the term anxiety has been used by Mowrer (9) and Miller (8) in theoretical interpretations of learning phenomena to mean an internal response conditioned to previously neutral cues which have been associated with noxious stimulation.

Recent animal experiments (1, 5, 8) have shown that the operations defining anxiety in learning theory studies produce a state which exhibits both the reinforcing and energizing properties of an acquired drive. Evidence has been obtained by Welch and Kubis (12) and by Taylor (11) with human Ss indicating that clinically defined anxiety may also exhibit these motivational properties. If, as there is reason to believe, the operations involved in either definition result in functionally equivalent drive states, then predictions involving the assumed relationship of the drive variable to other psychological variables should be equally applicable to both experimentally induced and clinically defined anxiety.

It has been previously reported (10) that, following Hull's postulates concerning the theoretical relationship between drive and effective habit strength, increases in experimentally induced anxiety produce a general elevation in the stimulus-generalization gradients of response amplitude and frequency

obtained from human Ss. The effects of two levels of clinical anxiety upon gradients of generalization will be considered in the present paper.

METHOD

Experimental design. Seventy-two male psychiatric patients from the Veterans Administration Hospital at Coatesville, Pennsylvania, and 72 male students from the introductory psychology course at the State University of Iowa served as Ss for the experiment. In the first part of the experiment, 36 patients were designated as the high-anxiety group and 36 patients as the low-anxiety group on the basis of psychiatric ratings. The 36 patients in each group were then randomly assigned to one of three experimental conditions: (a) strong shock, (b) weak shock, or (c) buzzer. Thus, there was a high-anxiety and a low-anxiety subgroup of 12 patients for each of these conditions, allowing an independent evaluation of the effects of the two levels of clinical anxiety, of the three experimental anxiety conditions, and of their interaction.

The second part of the experiment employed the 72 students in an exact replication of the above design. Anxiety level was determined, however, by means of a scale (11) which had previously been used to discriminate high and low anxiety in a normal college population. The replication served to determine whether clinical anxiety would have the same relative effects in both pathological and normal Ss, as well as providing a basis for the comparison of results obtained by the use of conventional psychiatric ratings of anxiety with those obtained from a more objective definition of anxiety.

Subjects. The selection of Ss with high clinical anxiety from among the hospitalized patients was accomplished by means of a simple scale for rating overt evidences of anxiety, developed in collaboration with the psychiatric staff doing the rating. Patients who showed no indications of manifest anxiety were matched for psychiatric diagnosis with individuals in the high-anxiety group so that differences obtained between these groups might not be attributable to pathological variations other than anxiety among these Ss. Of the patients selected, 40 were diagnosed as schizophrenic, 22 as psychoneurotic, and 10 as psychopathic personalities. The mean age of the high-anxiety group was 30.2; that for the low-anxiety group was 34.4. The mean IQ's for the two groups were 108.8 and 104.0, respectively. Neither of these differences between the groups was statistically significant.

Clinical anxiety level in the college population was defined by means of a modified form of the Anxiety scale described by Taylor (11). Fifty of the MMPI items which had shown the highest correlation with total anxiety score on the original scale were used in the

¹ This study was part of a doctoral dissertation completed in 1950 in the Department of Psychology at the State University of Iowa, under the direction of Drs. J. S. Brown and I. E. Farber. The author regrets the delay in publication and wishes to acknowledge those investigators (2, 6, 13) who have utilized the unpublished manuscript as a point of departure for further studies.

TABLE 1

MEAN AMOUNTS OF GENERALIZATION OF RESPONSE AMPLITUDE

Group	Normal Ss		Psychiatric Ss	
	High anxiety	Low anxiety	High anxiety	Low anxiety
Strong shock	22.71	14.74	18.89	12.98
Weak shock	11.39	12.59	12.50	10.91
Buzzer	12.80	10.86	11.08	11.70
All conditions	15.63	12.73	14.16	11.86

modified scale. In addition, the *F*, *K*, and *L* scales of the MMPI were included to detect false high- or low-anxiety scores. Any one of the following criteria served to invalidate an anxiety score: (a) an *F* score of 12 or above, (b) a *K* score of 24 or above, or (c) an *L* score of 7 or above. The new version of the test was administered to 574 introductory psychology students, and the obtained distribution of scores was almost identical to the distribution obtained by Taylor. Thirty-six male Ss from the upper 20 per cent of the distribution and 36 males from the lower 20 per cent comprised the high- and low-anxiety groups, respectively.

Procedure. The apparatus employed in the experiment and the details of the procedure have been described previously (10). In general, the procedure involved training Ss to make a simple motor reaction as quickly as possible upon the presentation of a rectangularly shaped visual figure. Three other rectangles, differing from the training stimulus in height (in steps of 0.25 inches), served as the generalized stimuli to which the Ss were instructed *not* to respond. Following 20 trials with the training stimulus, Ss were given five test trials on each of the four stimuli. These 20 test trials were interspersed among 60 additional training trials, so that each test trial was followed by three presentations of the training stimulus. The frequency, amplitude, and latency of "false" motor responses to the test stimuli were obtained by means of a movement-recording pulley system and served as the measures of generalization. Variations in the level of experimentally induced anxiety were produced by instructing Ss that slow responses would elicit either a strong shock, a weak shock, or the sound of a buzzer. Shocks and buzzes were presented during the instruction period and on the sixth and eighteenth training trials.

RESULTS

The major results of the experiment are presented in Tables 1 and 2, which show the mean amounts of generalization of response amplitude and the percentage frequencies for each of the six subgroups of normal and psychiatric Ss. Each of the means in Table 1 is based upon an over-all generalization score obtained for every S by summing the mean amplitude of his responses to the four stimuli on the 20 test trials; the means in Table 2 were similarly obtained from the total fre-

quency of responses during the 20 test trials, converted to percentages. Thus, these sum scores computed for each S represent essentially the heights of the individual generalization gradients, and their means designate the heights of the subgroup gradients. Latency data are not presented, since generalization gradients comparable to those found with the two principal measures were not obtained.

Main effects of clinical and experimental anxiety. The principal hypothesis of the present study was that high levels of clinical anxiety, as well as experimentally induced increases in anxiety, tend to elevate gradients of generalization. The Ss in the high-anxiety subgroups should therefore show greater amounts of generalization than those in the low-anxiety groups. Examination of Tables 1 and 2 reveals that the high-anxiety Ss tended to show more generalization of amplitude and percentage frequency in both normal and psychiatric groups. Statistical analyses of these data, a sample of which is presented as Table 3, indicates further that the difference between the two levels of clinical anxiety in amplitude of response is significant at the .05 level in the normal groups.

The difference between the two clinical-

TABLE 2

MEAN AMOUNTS OF GENERALIZATION OF RESPONSE FREQUENCY IN PERCENTAGE

Group	Normal Ss		Psychiatric Ss	
	High anxiety	Low anxiety	High anxiety	Low anxiety
Strong shock	73.75	57.10	65.00	50.85
Weak shock	39.60	46.25	47.50	42.90
Buzzer	42.90	38.75	40.05	48.35
All conditions	52.08	47.37	50.83	47.37

TABLE 3

ANALYSIS OF VARIANCE OF AMPLITUDE DATA FOR NORMAL SUBJECTS

Source of Variation	df	Mean Square	F
Clinical anxiety	1	151.67	3.99*
Experimental anxiety	2	371.77	9.78**
Clinical \times experimental	2	130.43	3.43*
Within groups	66	38.01	
Total	71		

* $p = < .05$ ** $p = < .01$

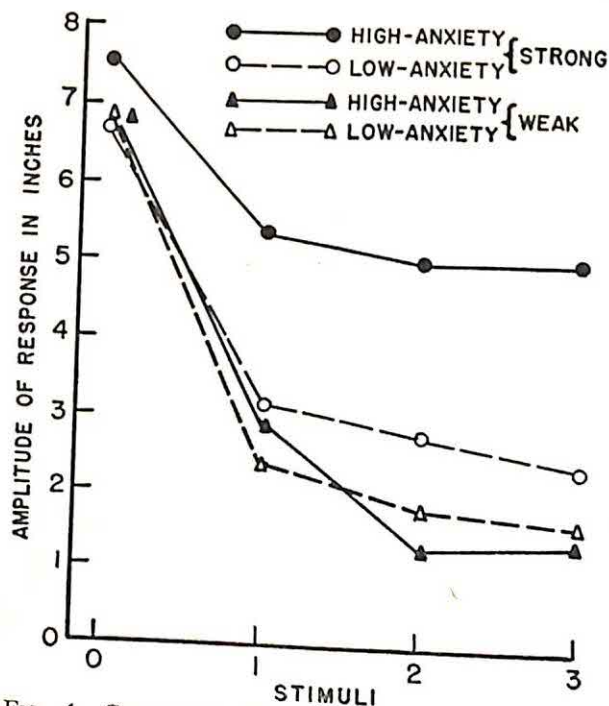


FIG. 1. GRADIENTS OF RESPONSE AMPLITUDE FOR HIGH- AND LOW-CLINICAL ANXIETY SUB-GROUPS IN THE NORMAL GROUP

The training stimulus is designated as 0; and the generalized stimuli are designated 1, 2, 3, in order of decreasing similarity to the training stimulus. Weak-shock and buzzer conditions have been combined in the curves labeled weak.

anxiety levels in generalization of amplitude, however, was not significant in the psychiatric Ss. Furthermore, the clinical anxiety variable produced no significant difference in mean frequency scores in either the normal or psychiatric Ss. In contrast to these generally negative findings for the main effects of clinical anxiety, experimental anxiety produced significant increases in generalization of amplitude and frequency for both normal and psychiatric Ss. As is apparent from Tables 1 and 2, however, while the strong-shock condition resulted in significant increases in generalization, the differences between the weak-shock and buzzer condition were not statistically significant.

Interaction of clinical and experimental anxiety. Further examination of Table 3 indicates that the interaction between clinical and experimental anxiety was significant at the .05 level. Careful scrutiny of Tables 1 and 2 and additional statistical comparisons reveal that this interaction effect is attributable to the fact that the high-clinical-anxiety groups exhibited significantly more generalization

than the low-anxiety groups *only* under the strong-shock condition.

This effect is presented graphically in Fig. 1, which shows the gradients of response amplitude obtained from the normal Ss under different experimental conditions. The weak-shock and buzzer groups have been combined in the curves labeled "Weak" for purposes of graphic presentation, since the differences between them were not significant. From Fig. 1 it is clear that strong shock consistently elevates the generalization gradient, but high clinical anxiety produced more generalization than low anxiety only when combined with strong shock. No significant difference in the height of the over-all generalization gradients was obtained between the high- and low-anxiety subgroups for either the weak-shock or buzzer conditions. Statistically significant interactions, essentially paralleling the results shown in Fig. 1, were obtained for frequency and amplitude measures in both the psychiatric and normal Ss.

DISCUSSION

It has been demonstrated in the present study that while experimentally induced increases in anxiety act as a drive in elevating gradients of generalization, high levels of clinical anxiety exhibited drive properties only when combined with strong shock. These results suggest that clinical anxiety may act to increase generalization only when the stimulus situation involves the threat of strong punishment. This finding is only partly confirmatory of the hypothesis that manifest anxiety symptoms indicate the presence of a strong anxiety drive, since the effects of this additional drive might also be expected to increase generalization under conditions of weaker experimentally induced anxiety.

One possible interpretation of this inconsistency is that clinical anxiety symptoms, as defined either by psychiatric ratings or scores on an anxiety scale, do indicate the presence of a drive-like anxiety state, but that the motivational properties of this state are not activated unless the conditioned anxiety response is elicited by appropriate cues or by sufficiently noxious stimulation. Cameron has described the chronic anxiety reaction as "characterized by the presence of persistently heightened skeletal and visceral tension . . . which predispose the individual to give exaggerated and in-

appropriate responses on relatively slight provocation" (3, p. 249). This description parallels the hypothesis that anxiety-produced increases in drives should elevate the generalization gradients of frequency and amplitude, since elevation of frequency would produce inappropriate responses and elevated amplitude would produce exaggerated responses to stimuli that are normally ineffective in eliciting responses from less anxious people. Thus, the results obtained suggest that clinical anxiety, as measured in the present investigation, may be equated with other drives as a state-variable within the organism only under the specified conditions in which particular anxiety-producing cues are present in the situation.

An alternative interpretation of the fact that no difference was found between the two degrees of clinical anxiety under weak shock conditions is that the criteria employed for determining the levels of clinical anxiety were inadequate. This possibility hardly seems likely, however, in view of the difference obtained under the strong-shock condition. Furthermore, the fact that both the psychiatric ratings and the anxiety scale produced similar results indicates that this interpretation is insufficient. In fact, the similarity of the results obtained by the use of these two techniques for identifying clinical anxiety suggests that they tended to identify the same process and thus contributes to the construct validity of the anxiety scale as a measure of symptomatic anxiety reflecting an acquired drive.

It may be of additional interest to note that the present findings agree only in part with those of several recent studies. While Wenar (13) found that both clinical anxiety and stimulus intensity increased generalization to temporal stimuli, he obtained differences between his high- and low-anxiety groups under weak-shock and buzzer conditions, as well as under strong shock. Since either a shock or a buzzer was the signal to react on every trial, and since Ss were always notified of their reaction time in Wenar's procedure, it would appear that these conditions are sufficiently anxiety arousing to elicit differences between clinical anxiety levels that were not obtained under the less threatening conditions of the present experiment.

A more serious discrepancy with the present

findings and also with those of Wenar, however, is a suggestion by Eriksen (4) that his results indicate that high-anxious Ss exhibit *less* generalization than low-anxious Ss. This inference is based, in turn, on his findings that hysteric Ss showed more generalization than psychasthenic Ss, and that there was a significant negative correlation between his hysteria-minus-psychasthenia dimension and the manifest anxiety scale. Since Eriksen's suggestion is the direct opposite of the present findings, those of Wenar, and those reported by Hilgard, Jones, and Kaplan (7) that high-anxious Ss give more conditioned eyelid responses to a generalized stimulus, it seems likely that the discrepancy suggested by his results may have been an artifact of the range of anxiety scores included in his experimental sample. It would also appear reasonable to note that the greater generalization shown by his hysteric Ss may well have resulted from their repressive tendencies producing greater inattention and consequently more false responses. If such a relationship prevailed between general responsiveness and hysteria, the positive relationship between anxiety and generalization, demonstrated both for clinical ratings and the manifest anxiety scale in the present experiment, might well have been obscured and even reversed.

SUMMARY

The present study was concerned with the effects of anxiety upon stimulus generalization. It was hypothesized that the state inferred from manifest clinical symptoms of anxiety would show functionally similar motivational properties to the state of anxiety defined in terms of an implicit response that has been conditioned to situations involving noxious stimulation. It was predicted that both types of anxiety would exhibit the energizing properties of a drive and therefore elevate response gradients of generalization.

Seventy-two male psychiatric patients and an equal number of college students were divided into high-clinical-anxiety and low-clinical-anxiety subgroups by means of psychiatric ratings and a scale designed to measure anxiety in students. Variations in experimental anxiety were effected within each group by the use of two different intensities of shock and a buzzer. The Ss were trained to make a motor

reaction to a visual stimulus and then tested on other stimuli differing from the training stimulus in height. The amplitude and frequency of responses to these generalized stimuli were employed as the measures of generalization.

The results showed that the groups designated as high in clinical anxiety showed significantly more generalization than the low-clinical anxiety groups under the strong-shock condition. No difference was found between the two levels of clinical anxiety for either the weak-shock or buzzer condition.

Markedly parallel results were obtained by the use of psychiatric ratings for differentiating the degree of clinical anxiety in the psychiatric patients and by the use of a psychometric scale with the normal Ss.

It was concluded that the heightened responsiveness of anxious individuals to a wide range of stimuli may be subsumed under the more general effects of increased drive upon stimulus generalization, but that clinically defined anxiety is activated as a drive variable only when certain noxious cues are present.

REFERENCES

1. BROWN, J. S., & JACOBS, A. The role of fear in the motivation and acquisition of responses. *J. exp. Psychol.*, 1949, **39**, 747-759.
2. BUSS, A. H. Stimulus generalization as a function

- of clinical anxiety and direction of generalization. *J. abnorm. soc. Psychol.*, 1955, **50**, 271-273.
3. CAMERON, N. *The psychology of behavior disorders*. New York: Houghton Mifflin, 1947.
4. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 561-565.
5. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111-131.
6. FARBER, I. E., & SPENCE, K. W. Relation of stimulus intensity, practice, anxiety, and sex to simple RT and temporal generalization. Iowa City: State Univer. of Iowa, June 27, 1955. (Tech. Rep. No. 4, Contract N9 ONR-93802.)
7. HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, **42**, 94-99.
8. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89-101.
9. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553-565.
10. ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35-43.
11. TAYLOR, JANET. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
12. WELCH, L., & KUBIS, J. Conditioned PGR in states of pathological anxiety. *J. Psychol.*, 1947, **23**, 83-91.
13. WENAR, C. Reaction time as a function of manifest anxiety and stimulus intensity. *J. abnorm. soc. Psychol.*, 1954, **49**, 335-340.

Received November 18, 1955.

DISTRACTIBILITY IN THE CONCEPTUAL PERFORMANCE OF SCHIZOPHRENICS¹

LOREN J. CHAPMAN

Northwestern University²

MANY studies have been carried out in which the performance of normal and schizophrenic Ss on conceptual tasks have been compared. Despite a variety of techniques, nearly all such investigations have found that the schizophrenic group performs in an inferior manner to a normal group. Various writers disagree, however, on the cause of the inferior performance. One point of view, espoused by Goldstein and Scheerer (3), Hanfmann and Kasanin (4), and others is that schizophrenics are deficient in ability to form concepts. Other workers, among whom Cameron (1, 2) has been outstanding, believe that the schizophrenics do poorly, not because of inability to form concepts per se, but because of behavior tendencies which interfere with adequate performance. Cameron believes that schizophrenics are overly responsive to distractions provided by both external and internal stimuli, and in responding to such stimuli, produce responses which are incompatible with correct solution. Thus, a schizophrenic may be able to form the necessary concepts in a test, but fail because response to distracters prevents the demonstration of his concept-forming ability by a correct performance.

This latter interpretation suggests a possible experimental arrangement for providing evidence concerning the role of interfering factors in schizophrenic conceptual performance. If instead of requiring *S* to form a concept, *E* pointed out to *S* the correct concept and asked *S* to use it to guide his performance, schizophrenics might be expected to become increasingly inferior to normal Ss in following these instructions as more and

more distracting elements were introduced into the context of the situation. Assuming that the required correct solutions were the same throughout all degrees of distraction, such a differential rise in errors for the two groups could not be predicted from an inability of the schizophrenics to form concepts, since only the context of the task and not the task itself would be modified.

This distraction hypothesis could be tested in a number of ways, the simplest perhaps being variation of the number of distracters.

The dimension actually used in this study was a modification of this procedure, involving a card-sorting task in which *S* was required to assign several series of response cards (here called RCs), each bearing four figures, with one of three standard cards (SCs), each one also with four figures on it. Fig. 1 shows some examples of the cards. The *E* asked *S* to sort each RC on the basis of some commonality with an SC, but in doing so to use only the lower right-hand corner of the SC, ignoring the other three figures. (By making the correct sorting basis always the lower right-hand corner, *E* minimized the importance of memory in following the instructions.) Each RC contained one figure that enabled it to be correctly placed with one SC. Its other three figures, however, shared varying numbers of commonalities with figures on a second SC that were not in the lower right-hand corner and which, therefore, were incorrect for sorting. There were two kinds of RCs, namely identity RCs and concept RCs. On an identity RC, of which RC 24 in Fig. 1 is an example, one figure was identical to an SC figure that was correct for sorting (in the lower right-hand corner of the SC) and had other figures that were identical to figures on a second SC, these SC figures being incorrect for sorting (not in the lower right-hand corner of the SC). For example, RC 24 has an octopus which is identical to the octopus in the lower right-hand corner of SC I, and also has an apple and a hammer, identical to figures in positions incorrect for sorting on SC III. The letter *Z*

¹This paper is based on a Ph.D. dissertation at Northwestern University. The manuscript in its present form was written while the author was a Public Health Service Research Fellow of the National Institute of Mental Health. The author is indebted to Dr. Janet A. Taylor for invaluable assistance and advice in directing the dissertation, and to both Dr. Taylor and Dr. Howard F. Hunt for their critical reading of the manuscript.

²Now at the University of Chicago.

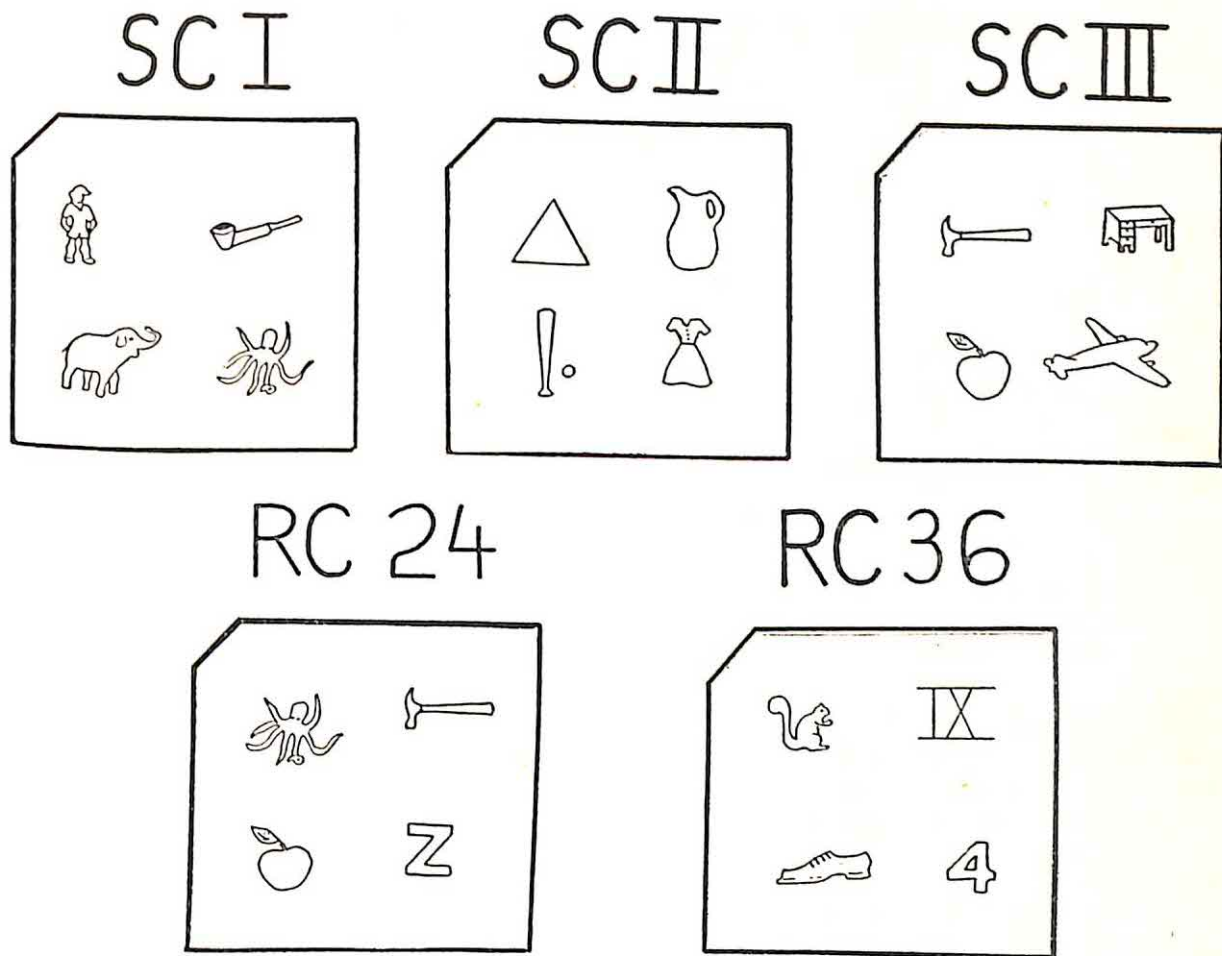


FIG. 1. EXAMPLES OF THE STANDARD CARDS AND RESPONSE CARDS

on RC 24 is a "filler" figure, designed to share no commonalities with SC figures but used on those RCs with less than three distracter figures to keep the number of such figures for each RC at four.

A concept RC, of which RC 36 is an example, differs from an identity RC only in that the commonalities both for correct sorting and for distracters are conceptual. For example, RC 36 has a shoe which shares a concept with the dress which is in the position correct for sorting on SC II, and also has a squirrel which shares a concept with the elephant, which is in a position incorrect for sorting on SC I. The Roman numeral IX and Arabic number 4 are filler figures on RC 36.

The number of commonalities present which were incorrect for sorting, here called the distracter level, was varied from 0 to 3. The prediction was that the schizophrenics would respond much more than normal Ss to these distracting commonalities. No prediction

could be made concerning a difference between the two groups at the zero distracter level because the task might be such a simple one (both in difficulty of concepts and in terms of distracters) that schizophrenics might do as well as normal Ss. Rather, it was predicted that regardless of the comparative level of performance of the groups initially, the errors of the schizophrenics would rise with an increase in distracters while the errors of the normal Ss would not, producing a divergence in the error curves.

One reason for making the choice of manipulating the number of commonalities rather than sheer number of figures was the greater meaningfulness that the use of this latter dimension would give to positive results. If it were found that schizophrenics but not normal Ss make increasing errors with an increase in the number of distracters involving conceptual commonalities, it would give further evidence that the schizophrenic de-

fect in conceptual performance is not wholly one of concept forming; before a response to conceptual distracters can deter performance, *S* must obviously formulate a concept.

METHOD

Subjects. The final experimental groups, after pre-testing, consisted of 25 normal and 40 schizophrenic *Ss*. The normal *Ss* were 21 Naval Corps School students at Great Lakes Naval Training Station and 4 Downey VA Hospital aides.

Schizophrenic *Ss* were chosen from those patients at the Downey Hospital bearing the diagnosis of chronic schizophrenia whose case material did not indicate either cerebral organic involvement or mental deficiency. Other prerequisites were that the first hospitalization for psychosis had occurred at least one year previously, that the most recent hospitalization had occurred at least 3 months previously, that the patient had not received electric shock therapy within the past 3 months and had never received more than 35 treatments. No selection was attempted on the basis of diagnostic subtype.

All *Ss* were males between 18 and 51 years of age and had between 7 and 16 years of education. The mean age, education, and functioning IQ for the schizophrenic group were 29.6, 11.4, and 95.1. The corresponding values for the normal group were 23.7, 10.6, and 99.9. The IQ scores were obtained from the CVS for the schizophrenic *Ss* and hospital aides, and from GCT equivalents for the 21 naval personnel. Results of *t* tests and *F* tests showed that the groups did not differ significantly on mean or variability in either education or IQ. On age, the schizophrenics were found to have a higher mean and a lower variability. However, there seems to be no reason to suspect that the age variable is correlated with response to the experimental variable being tested, so that the discrepancy between groups was not considered to be important.³

Experimental task. The task materials consisted of a number of cards, each $4\frac{1}{2}$ inches square with four figures drawn on it in India ink. There was a total of 60 different figures used on the cards, the figures being chosen so as to fall into 15 conceptual categories of four figures each. With the help of a preliminary study, the conceptual categories and their member figures were chosen with a view toward keeping the individual figures recognizable and the concepts relatively non-overlapping and simple.

³Matching of groups on any characteristic presupposes, of course, that the characteristic is correlated with the variable under experimentation. In the present study, since the experimental dimension was a new one, the only choice was to match on such obvious (but not necessarily related) factors as IQ and education. The discrepancy in age was not deemed relevant. A check on the data of this and a later experiment by the writer revealed no relation between response to distracters and age for either the normal or schizophrenic *Ss*. In fact, for both groups of the present study, the younger *Ss* actually made a few more errors, though the difference was not significant. Greater errors on the part of the younger *Ss* would, if present, reduce the likelihood of obtaining the predicted difference between groups.

Of the 15 categories, 12 were used as sorting categories, i.e., used as concepts according to which *Ss* were expected to sort. These sorting categories were smoking equipment, clothing, geometric figures, sports equipment, furniture, means of transportation, tools, fruit, animals, sea life, people, and kitchen utensils. The other three categories, used as "filler" figures and never employed as sorting concepts but placed on some of the response cards to keep constant the number of figures at four, consisted of letters, Arabic numerals, and Roman numerals.

For each sorting task, a set of three SCs was used. For all sortings, the same 12 figures were used on the SCs, each of the figures being from a different category. Four different sets of SCs were obtained by moving the four figures on each of the three cards one position clockwise. One of the four figures on each SC, always the lower right-hand corner figure, was arbitrarily designated as the "correct" basis of sorting a particular series of RCs.

In addition to a correct and a distracter SC for each identity or concept RC, there was a third SC containing no figures that shared identities or concepts with it. This was called the irrelevant SC, and provided an index of errors other than those resulting from sorting with distracter figures. Which one of the three SCs fulfilled each of these three relationships varied from RC to RC.

Apparatus. Three $5\frac{1}{2}$ × $8\frac{1}{2}$ " wooden boxes were placed on a table in front of *S*. On the vertical face of each box was a slot in which the RCs were to be deposited. On the top of each box there was a rack for holding the SC.

Preliminary training and screening. Before the experiment proper, two preliminary tasks were given both to pretrain *Ss* in the materials used and to identify those *Ss* unable to handle them. The *S* was shown the 60 figures used in the experimental task, each on a separate card, and asked their names. If *S* failed to give an approximately correct name he was told the answer. However, if *S* failed more than five figures on his first try (not including values of Roman numerals), or any one figure on a second presentation, he was dropped from the experiment. None of the normal group and six of the schizophrenic *Ss* were eliminated on this basis.

Next, *E* gave *S* a preliminary sorting task, again using the 60 single figure cards. The *E* placed four stimulus cards from different categories on the table and asked *S* to take cards from the top of the pile in order and place them with that stimulus card which pictured something of the same kind. When *S* had placed three additional cards with any stimulus card, *E* removed this category and replaced it with a new stimulus card that started a new category. The order of the cards in the pile was such that *S* was always able to place each card with one of the four choices confronting him. If necessary, *S* was helped or corrected, but if this happened on more than five cards, he was excluded from the experiment. None of the normal *Ss*, but five schizophrenic *Ss* were eliminated on this basis.

Presentation of experimental task. The *S* was instructed to choose the correct box for each RC by finding a figure in one of the four corners of the RC that was either identical to or of the same sort as the figure

in the lower right-hand corner of one of the three SCs. The *E* asked *S* to explain these instructions before he began sorting. Only those *Ss* were used who verbalized that they were to use the lower right-hand corner of one of the SCs and any corner of the RC. None of the normal *Ss*, but 19 of the schizophrenic *Ss* were dropped on this basis. The *S* was again asked to explain the instructions after his first series, and also after his last. On one or the other of these two repetitions, 8 of the schizophrenics failed, but they were not excluded from the experiment since they had all adequately explained the instructions on the first request before beginning to sort.

Each *S* received one 36-card series from each of the four distracter levels, each such series being divided into a subseries of 18 concept RCs and a subseries of 18 identity RCs.

Order of presentation of the series representing the four distracter levels was counterbalanced among the *Ss*. Appropriate counterbalancing was also used to prevent biasing of results by preference for any of the box positions, figure position on the RCs, any of the various concepts and figures, or any differential difficulty in using the above, or by practice and fatigue effects.

RESULTS

Since the data were extremely skewed, nonparametric statistical techniques were used in all analyses.

Four types of error scores were used. The number of incorrect placements of the RC with the SC containing the distracter figures, Distracter Score, is a crude measure of sorting in response to distracters. The number of placements with the Irrelevant SC, Irrelevant Score, measures errors which are not in response to the distracter figures. These errors include mistakes due to carelessness, non-cooperation, and sorting by commonalities which *S* may have perceived but which *E* did not intend to build into the test. Because, in any given subseries, errors of the irrelevant type should be made equally often in response to each of the three SCs, the difference between the two error scores, Distracter-minus-Irrelevant Score, is considered a purer measure than Distracter Score alone of sorting with the distracter figures. The final measure, Total Errors, is a measure of errors of all types. It is the only error score which can be used at the zero distracter level because without distracters there is obviously no Distracter Score.

Most of the analyses were done separately for the concept RCs (in which both the commonalities correct for sorting and the distracter commonalities were concepts), and

TABLE 1
MEAN DISTRACTER AND IRRELEVANT SCORES FOR SCHIZOPHRENIC AND NORMAL *Ss* ON CONCEPT AND IDENTITY RESPONSE CARDS (RCs) AT EACH DISTRACTER LEVEL

Distracter Level		Distracter Score		Irrelevant Score	
		Schiz.	Normal	Schiz.	Normal
0	Concept RCs			.98	.20
	Identity RCs			.72	.12
1	Concept RCs	2.15	.40	.75	.20
	Identity RCs	2.00	.48	.48	.16
2	Concept RCs	3.65	.68	.80	.08
	Identity RCs	3.87	.44	.48	.00
3	Concept RCs	3.22	.68	.55	.00
	Identity RCs	2.60	.40	.42	.12

for the identity RCs (in which both the correct and the distracter commonalities were identities).

*Performance of the normal *Ss*.* The normal group as a whole made very few errors. For the total of 144 card placements which each *S* made in the experimental task, 21 of the 25 *Ss* made 2 or fewer errors. The remaining 4 *Ss* erred, respectively, 3, 5, 7, and 65 times. The distribution of the Distracter and Irrelevant errors among the distracter levels for both concept and identity RCs is shown in Table 1, and the relationship between distracter level and Total Errors is shown in Fig. 2. Although there is a hint of an increase in Total Errors

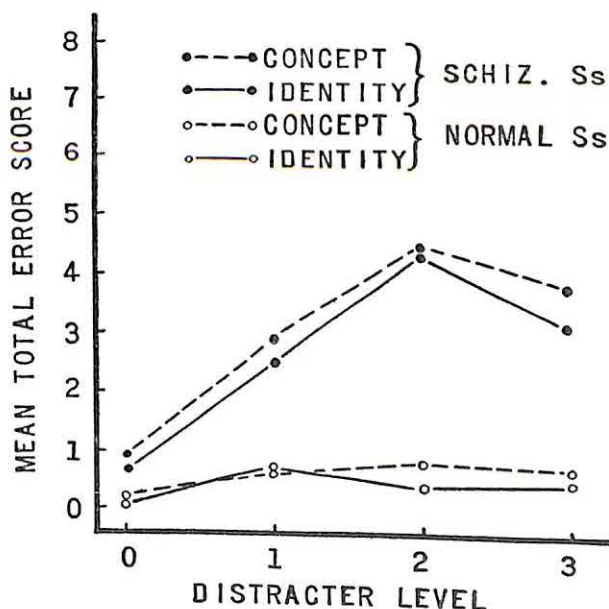


FIG. 2. RELATION OF TOTAL (T) ERROR SCORE TO DISTRACTER LEVEL

with the distracter level for these normal Ss, due primarily to the single deviate normal *S* who made 65 errors, analysis of variance by ranks (5) showed for each of the two types of RCs no over-all difference on Total Errors among distracter levels. A similar lack of relationship was found between the other error measures and distracter level.

Performance of the schizophrenic Ss. Of the 40 patients, 18 made 3 or fewer errors, while the remaining 22 ranged as high as 108 errors in their 144 card-placements. As seen in Fig. 1, there is a strong relation between Total Error Score and distracter level, with the largest error difference occurring between distracter levels 0 and 1. Analyses of variance by ranks (5) showed an over-all significant difference between the four levels, both for concept RCs ($p < .001$) and for identity RCs ($p < .01$). In both cases, the rise was significant by a Wilcoxon test from 0 to 1 and from 1 to 2, but level 3 did not differ significantly from either 1 or 2. Thus, in general, the hypothesis is confirmed that as the number of distracters increases, the number of errors increases for schizophrenics but not for normal Ss.

For the Distracter and Distracter-minus-Irrelevant Scores, inspection of Table 1 suggests a similar tendency for errors to rise from level 1 to 2 and a decrease, smaller in size, from 2 to 3. An analysis of variance by ranks, however, did not show an over-all difference among the three distracter levels. This lack of significance is not too surprising when one considers that a large part of the rise in errors with distracters is, as shown in Fig. 2, between the 0 and 1 distracter levels. Since the Distracter Score cannot be obtained at the 0-distracter level, the rise from 0 to 1 must be omitted from analysis on these two error measures, thus reducing the probability of obtaining over-all differences.

The data were examined to determine whether Total Error Scores increase within each 18-card subseries as a function of time since instructions, a finding which would be expected if the response to distracters resulted from a sheer forgetting of the instructions. The position of the card in the order of presentation was used as an approximate measure of time. When mean number of errors for each successive block of three cards was

plotted, there was no progressive increase in errors within a subseries. This observation, together with the fact that only 8 of the Ss failed to explain adequately the instructions on either the second or third request, indicates that forgetting of the instructions is of little or no importance in accounting for the schizophrenics' response to the distracters.

It will be remembered that many of the schizophrenic Ss performed the experimental tasks with very few errors, equalling the performance of the normal Ss in this respect. A further analysis of the schizophrenics' data was made to determine whether one of the factors contributing to response to distracters by a schizophrenic *S* was severity of illness. The patients were therefore divided into two groups, 23 patients given full privileges and 17 locked-ward patients. This division was used as a rough measure of severity of illness. They were also divided into 18 high scorers on Total Errors (12 or more errors on concept and identity RCs combined), and 22 low, and into 17 high Distracter - minus - Irrelevant scorers (score of 4 or more on concept and identity RCs combined), and 23 low. This latter grouping is of special interest here because it is based on errors independent of any difficulty which *S* may have had in handling the materials conceptually. Chi square showed that the locked-ward patients included more of the high Total Error scorers ($p < .05$) and of the high Distracter-minus-Irrelevant scorers ($p < .02$). Thus, even though the present experiment included only relatively undisturbed patients (the most disturbed patients in the hospital could not meet criteria for inclusion), as degree of schizophrenic pathology increased, so did response to distracters.

Comparison of normal and schizophrenic Ss. As has already been discussed, the performance of the schizophrenics was adversely affected by an increase in the number of distracters, while the performance of the normal Ss remained constant. Turning to a more direct comparison of the performance of the two groups, the schizophrenics exceeded the normal Ss on Total Error Score for the four distracter levels combined both on concept RCs ($p < .01$) and on identity RCs ($p < .01$), significance being tested by a Mann-Whitney *U* test. More enlightening is the comparison of the two groups at the separate distracter

levels. As seen in Fig. 1, at the zero distracter level, the schizophrenics made only a few more errors than the normal Ss, and the difference fell short of significance both on concept RCs ($p = .44$), and on identity RCs ($p = .46$). At the 1-, 2-, and 3-distracter levels, however, the differences were much greater, and in each case, for the concept RCs, were significant at the .01 level or less.

Thus, the presence or absence of distracters seems to be an important determining factor in whether or not the schizophrenics do more poorly than normal Ss on a conceptual task.

On the Distracter Score, and the Distracter-minus-Irrelevant Score, the schizophrenics also exceeded the normal Ss for the three levels combined (the zero-distracter level being necessarily omitted), both on concept and identity RCs ($p < .02$ or less), though for some of the levels taken separately the p values fell short of the .05 level of significance. The significant difference on the over-all Distracter-minus-Irrelevant Score for concept RCs is of special importance. This score measures errors in performance independent of errors arising from sources other than response to distracters; patients unable to grasp concepts, therefore, would be expected by chance to have Distracter-minus-Irrelevant Scores clustering around zero. Consequently, the greater error scores of the patients on this measure are clearly due to response to the distracters rather than to a lowered ability to sort in accordance with concepts.

DISCUSSION

The finding that the schizophrenics' errors rose as distracter level increased while those for normals did not confirms the hypothesis that much of the inferior conceptual performance of schizophrenics may arise from a response to distracters. This response may itself entail conceptual behavior as is borne out by the fact that the schizophrenics were affected adversely by the conceptual type of distracter as well as by the identity type. However, these results do not show whether or not there is also present in schizophrenics some lowered ability to form concepts. Not only do patients labeled schizophrenic differ a great deal among themselves in the kind of aberrant solution offered on conceptual tests,

but individual schizophrenics often show a variety of such solutions. One must be cautious in elaborating any single finding to explain all such aberrations. Future research should indicate the importance of response to various kinds of distracters, as well as reveal other sources of schizophrenics' inferior conceptual performance.

SUMMARY

The hypothesis tested was that the inferior performance of schizophrenics on conceptual tasks is at least in part due to their handling the test materials in other ways than those required by instructions, rather than due only to an inability to form concepts. Normal and schizophrenic Ss were asked to sort response cards containing four figures with three four-figured standard cards, one figure on each standard card being arbitrarily designated by instructions as correct for sorting with the rest incorrect. Each response card contained one figure that shared some commonality with the correct figure on one standard card. Also, the figures on each response card were varied so that from zero to three of them were identical to or shared concepts with incorrect figures on a second standard card. The schizophrenic group much more than the normal Ss used the incorrect "distracter" commonalities as a basis for sorting, even when this required conceptual sorting. The error curves for the two groups therefore diverged as the number of distracter commonalities increased. The results are interpreted as supporting the hypothesis.

REFERENCES

1. CAMERON, N. Reasoning, regression and communication in schizophrenia. *Psychol. Monogr.*, 1938, 50, No. 1 (Whole No. 221).
2. CAMERON, N. Deterioration and regression in schizophrenic thinking. *J. abnorm. soc. Psychol.*, 1939, 34, 265-270.
3. GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monogr.*, 1941, 53, No. 2 (Whole No. 239).
4. HANFMANN, EUGENIA, & KASANIN, J. S. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr. Ser.*, 1942, No. 67.
5. WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Holt, 1953.

Received December 8, 1955.

SCALING NORM CONFLICTS IN THE AREA OF PREJUDICE AND DISCRIMINATION¹

NATHAN KOGAN AND JAMES F. DOWNEY

Harvard University

SINCE the publication of *The American Soldier* series, there has been a growing interest in the problem of multiple reference groups operating at cross purposes. Merton and Kitt (6) have been particularly forceful in calling for additional theory and research in this area, while others (10, 11) have designed ingenious questionnaires for tapping important aspects of norm and role conflict. Essentially, these investigators have sought to determine the manner in which discrepancies are resolved between divergent or contradictory norms prescribed by an individual's reference groups in a particular situation.

As of the present time, the norm-conflict model has not been employed in the context of prejudice and discrimination. Yet events on the American social scene suggest that such an application may yield dividends. In particular, incipient desegregation in public schools located in various parts of the South may result in conflicts between potentially tolerant attitudes toward Negroes and norms of intolerance fostered by parents, peers, and others in the community. Where there are attitudinal and behavioral changes toward minority group members in the direction of greater tolerance (or intolerance), the extent to which such change represents the resolution of a norm conflict could well bear investigation.

The present paper attempts the development of a scale measuring the propensities of individuals to choose a discriminatory, "hands-off," or militantly antidiscriminatory course of action in particular social situations. In the sense that the social situations are incidents depicting various kinds of prejudice and discrimination by parents, friends, and others, the choice of a course of action necessarily involves a distinctive resolution of a norm conflict.

METHOD

Subjects. The Ss were recruited from the high-school population of a New England city. For many years,

¹ The authors wish to thank Miss Basilla Neilan for her cooperation in making subjects available for this study.

this city experienced considerable intergroup conflict. Segregation of ethnic and religious groups resulted in few intergroup contacts and considerable intergroup hostility. Occasionally, adolescent "gangs" organized along racial, religious, or ethnic lines erupted into overt "warfare" with each other. The prevalent norms about minority subgroups were of a distinct ethnocentric variety.

With the support of certain community leaders, a young female social worker established a youth club (approximately two years before the date of this study) with the intention of harmoniously bringing together adolescents in the various community subgroups. By introducing athletic, social, and community participation programs, she has succeeded in lessening intergroup tensions, and, to a real extent, has established a group norm of antiprejudice militancy.

At the time the study was conducted, the total membership of this youth organization exceeded 1,400 boys and girls. A sample of 100 (50 boys and 50 girls) ranging in age from 12 to 20 was selected from (a) the club leadership, (b) members who appeared at club headquarters after school hours, and (c) members who appeared at a weekend social function. Except for the exclusion of Negro members, the sample is fairly representative of the club population. A range of Ss extending from leaders to persons on the periphery of the group is included in the sample.

The social situations questionnaire. The Ss filled out a questionnaire in which they were required to choose one of four alternative responses for resolving fifteen anti-Negro discrimination situations. These conflict situations include various kinds of discrimination ranging from opposition to intimate social equality, at the one extreme, to the threat of physical violence, at the other. The situations are classified into three clusters of five items each in terms of the person or persons manifesting the discriminatory behavior. One item from each of these three clusters—peers, authorities, and strangers—is presented here in the order cited.²

Imagine that you are planning to go out to a show with several friends and their dates. The day before you go out, one of the fellows calls you up and asks you if it would be all right if S (a Negro whom you do not know) could come along with his

² The complete questionnaire, item-by-item response frequencies, and derived Guttman scales have been deposited with the American Documentation Institute. Order Document No. 4885 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C. remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

date. You're pretty sure that the other kids in the party would not like this. Under these conditions,

—I would definitely say "No" or else get out of the party.

—I would probably say I would rather not, because of how the other kids would feel about it.

—I would probably say that it would be OK with me.

—I would definitely say that it would be OK and would try to convince the others that this was a good idea.

Imagine that as you are sitting at home one day, a neighbor comes in to ask your parents to sign a petition which would prevent Negroes from buying or renting land on your block. He explains that it would not hurt the Negroes because there are plenty of other good places in town to live. This move would not only save your section just for White people, but would also make your land worth more money. Your folks are just about to sign the petition. Under these conditions,

—I would insist that they were wrong and try to persuade them not to sign the petition.

—I would probably tell my parents that I didn't think they were doing the right thing.

—I would probably keep quiet because it wouldn't make much difference to me one way or the other.

—I would definitely approve.

Imagine that you are visiting a friend of yours out of town. You decide to go swimming at a local pool. Getting into line to pay your admission, you notice a young colored boy in line just in front of you. As he gets up to the booth to buy his ticket, the woman in the booth tells him that they don't allow Negroes in the pool. Under these conditions,

—I would complain to the woman about this policy.

—I would voice a complaint to my friend making sure that the woman would hear.

—I wouldn't say anything about it then, but might make a criticism to my friend later.

—I would say nothing, feeling it is the right of the pool owners to reject Negroes if they wish.

Questionnaire administration and scoring. The Ss specified their age, sex, religious preference, and extent of participation in the youth group, but did not sign their names to the questionnaire. The social worker in charge of the youth organization was not present during the test administration and the respondents were informed that she would not see the questionnaires. Item responses were scored by means of Guttman scaling procedures (3).

RESULTS

Guttman scales³ were obtained for the three item clusters and for all of the items taken

³In accordance with common scaling procedure, unambiguous imperfect scale patterns were assigned to perfect scale types in such a way as to minimize error. Where imperfect scale patterns were ambiguous, i.e.,

TABLE 1
PRODUCT-MOMENT INTERCORRELATIONS AMONG TOTAL
SCALE AND SUBSCALES

	Total	Peers	Author- ities	Strangers
Total	—	.77	.51	.52
Peers	.82	—	.47	.44
Authorities	.55	.52	—	.37
Strangers	.55	.48	.41	—

Note.—The correlation coefficients above the diagonal are uncorrected. Those below the diagonal have been corrected for the small number of categories by means of the Peters and VanVoorhis (8) formula. All correlations are significant beyond the .01 level.

together. The latter proved to be an 8-item scale with a coefficient of reproducibility of .91. The peers, authorities, and strangers items yield coefficients of reproducibility of .93, .96, and .95, respectively. All of the items in the peers and strangers clusters and all but one in the authorities cluster are represented in the subscales.

Despite high coefficients of reproducibility for the subscales, four or five dichotomized items are too few to guarantee unidimensionality. Guttman (4) maintains that the probability of obtaining a scale by chance is negligible for a sample of 100 on six dichotomized items, yet he questions the wisdom of generalizing this finding to an hypothetical item universe. While a reasonable degree of confidence attaches to the general scalability of norm conflict situations in the area of Negro-White relations, the positive findings respecting scalability within specific reference group clusters are necessarily of a tentative nature.

The intercorrelations among the scales are shown in Table 1. Differences in the extent of the correlation between the total scale and the subscales reflect the number of items contributed by each of the subscales to the total

could be assigned to more than one perfect scale type, the respondents were assigned to the scale type with the largest number of cases in accordance with Henry's (5) recommendations. Only those items meeting the "criterion of improvement" were included in the final scales.

In order to yield satisfactory Guttman scales, the four response categories had to be dichotomized as either 1 vs. 2, 3, 4 or 1, 2 vs. 3, 4 where 1 represents the most and 4 the least antidiscriminatory response. Very few Ss checked alternative 4, indicating that attitudes for the group as a whole are skewed toward the antidiscrimination pole.

scale. The peers subscale is the greatest contributor. The intercorrelations among the subscales are moderately high, suggesting that tendencies in the direction of discrimination or antidiscrimination cut across cluster areas.

To test the validity of the total scale, Ss were placed in one of four activity groups depending upon the extent of their *actual* participation in the youth club. Activity group assignments were made before the determination of members' scale types. Group I included the leaders, while Group IV consisted of members who did not actively participate in any of the group's activities. Groups II and III represented intermediate levels of participation. The product-moment r between activity group and total scale type is .36, a value statistically significant at the .01 level. When corrected for the limited number of categories by the Peters and VanVoorhis (8) formula, r becomes .40. Product-moment r 's between subscale types and activity group are .41 for peers, .44 for authorities, and .32 for strangers. Applying the same correction, the correlations become .47, .51, and .37, respectively. For each scale, Ss assigned to the more militantly antiprejudiced scale types tend to be the more active members of the group.

DISCUSSION

With the demonstrated scalability of the norm conflict situations, one can begin to ascertain just how far an individual will go in resisting discrimination stemming from a variety of sources. Scalability of the item clusters, when firmly established, should permit an even more refined estimate by ordering the "conflictful" quality of situations within specific reference groups. In the peers subscale, for example, a progressively increasing number of Ss endorses antidiscriminatory responses as item content extends from a disparaging joking remark about Negroes to the threat of physical violence.

The significant correlation between activity group and scale type supports the validity of the outlined instrument. The more active the group member, the stronger appears to be the incorporation of the youth-group norm.

The items developed in the present study are clearly most appropriate for adolescents who have some social contact with Negroes. However, where social class, geographical area, or

other factors produce conflicts differing from those embodied in the items of the present study, it would be advisable to devise new items and scale them accordingly. As a methodological contribution, the present investigation makes no claims respecting generalizability to other subject populations or to different sets of items, but merely purports to demonstrate the potential unidimensional scalability of norm-conflict situations in the intergroup relations area.

It should be noted that the whole gamut of attitudes and behaviors subsumed under the rubric of prejudice and discrimination can hardly be forced into a norm-conflict mold. As Allport (1) has indicated, prejudice can be conceived as a continuum extending from sheer conformity to social norms on the one extreme to a high degree of functional significance for the personality at the other pole. While the present authors strongly endorse this view, they would maintain that conformity pressures may assume heightened significance in an adolescent group. For many of the members of the youth club described above, ethnocentric loyalties fostered by parents and peers were transformed into personal loyalty to a strong group leader (the social worker) with militantly tolerant attitudes. White adolescents in other communities may, to a considerable extent, be caught between existent conformity pressures toward discrimination and increasing tolerance induced by widening social contact in school and out-of-school situations.

Others (7, 9) have devised "social situations" tests for measuring behavioral predispositions in a variety of areas. Items for these instruments were selected by the Thurstone equal-appearing intervals method. The intention of those investigators was one of bringing attitude scales a step closer to the behavioral level. Somewhat earlier, Bogardus (2) devised his Social Distance Scale, an instrument with latent unidimensional properties. The present scale attempts to embody the advantages of these earlier instruments, while incorporating recent advances in unidimensional scaling techniques and the theory of social norms and reference groups.

There are distinct limitations in the use of a social situations instrument. The direct quality of the items suggests that some Ss may respond in a deliberately self-enhancing manner,

whether this be in the direction of one or another set of norms. This problem, of course, is common to all attitude scales. Unlike attitude scales, however, the present instrument is explicitly intended as an indicator of prospective behavior. There is a distinct possibility that the checked alternative, though subjectively honest, is self-deceptive, for the S's behavior in an actual situation might not coincide with his response on paper. Finally, a general activity vs. passivity or interventionist vs. noninterventionist dimension may account for some of the variance in the scores. Some Ss would militantly follow the youth club's interests in any situation where the club clashed with persons or institutions external to it. The intercorrelations among the subscales are not so high, however, as to indicate that this militancy is an undifferentiated dimension.

Since the purpose of the present paper is essentially methodological, many substantive questions remain unanswered. While the potential unidimensional scalability of norm-conflict situations in the intergroup conflict area has been demonstrated, the reasons why specific situations are resolved in a specific direction for specific individuals have not been treated. Future theory and research, as Merton and Kitt (6) stress, should move on to consider "the dynamics of selection of reference groups among the individual's several membership groups."

SUMMARY

A sample of 100 high-school students—members of a youth club with a militantly antiprejudiced norm—filled out a questionnaire requiring them to choose one of four courses of action for resolving a series of Negro-White conflict situations. The questionnaire yielded an 8-item Guttman scale with a coefficient of reproducibility of .91. When the items were classified in one of three clusters in terms of the protagonist of the discriminatory act—peers, authorities, and strangers—these

clusters also yielded Guttman scales with high coefficients of reproducibility. The subscale findings were rendered tentative by the relatively small number of dichotomized items in each of the three clusters. Intercorrelations between the total scale and the subscales and among subscales were moderately high. Scale validity is supported by a significant correlation between scale type and extent of participation in the youth group. A note of caution is sounded respecting generalizability of the present findings to other subject populations and item universes. The present scale is compared with relevant antecedent instruments and the scale's limitations are discussed.

REFERENCES

1. ALLPORT, G. W. Prejudice: A problem in psychological and social causation. *J. soc. Issues, Suppl. Series*, 1950, No. 4.
2. BOGARDUS, E. S. *Immigration and race attitudes*. Boston: Heath, 1928.
3. GUTTMAN, L. The Cornell technique for scale and intensity analysis. *Educ. psychol. Measmt*, 1947, 7, 247-280.
4. GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer, et al., *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. Pp. 60-90.
5. HENRY, A. F. A method of classifying nonscale response patterns in a Guttman scale. *Publ. Opin. Quart.*, 1952, 16, 94-106.
6. MERTON, R. K., & KITT, ALICE S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the scope and method of "The American soldier"*. Glencoe, Ill.: Free Press, 1950.
7. PACE, C. R. A situations test to measure social-political-economic attitudes. *J. soc. Psychol.*, 1939, 10, 331-344.
8. PETERS, C. C., & VANVOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.
9. ROSANDER, A. C. An attitude scale based upon behavior situations. *J. soc. Psychol.*, 1937, 8, 3-15.
10. STOFFER, S. A. An analysis of conflicting social norms. *Amer. sociol. Rev.*, 1949, 14, 707-717.
11. STOFFER, S. A., & TOBY, J. Role conflict and personality. *Amer. J. Sociol.*, 1951, 56, 395-406.

Received December 12, 1955.

DEVELOPMENT AND EVALUATION OF A SCALE FOR MEASURING SOCIAL ACQUIESCENCE¹

BERNARD M. BASS

Louisiana State University

IN A preceding report (3), the suggestion was made that acquiescence, defined as the tendency to accept any generalization about behavior, is a parsimonious concept that accounts for a large percentage of the variance among many personality measurements. Acquiescence to a variety of other stimuli has also been observed (1, 3, 7, 8, 9).

The forms of acquiescent behavior tend to be fairly independent of each other (11, 14). The purpose of the present study was to develop and evaluate a measure of social acquiescence—acquiescence to a wide variety of generalizations concerning how persons behave or should behave. In agreement with others (2, 5, 6, 7, 9, 11, 12), it was believed that developing such a measure should provide a useful personality assessment device for the social psychologist.

METHOD

Initial Subjects and Pool of Items

An initial sample of 200 college students, half from the South and half from the Midwest, indicated whether they agreed with each of 300 statements. Fifty-five per cent of the sample was male. The mean age of the subjects (Ss) was 21.0 years ($SD = 3.5$); the mean education was 14.1 ($SD = 1.5$).²

The Ss were directed to "indicate whether you agree, disagree or are uncertain about" each of 300 famous sayings. The statements (maxims, aphorisms, adages, proverbs, apothegms, and sayings) concerned thirteen areas of human behavior. Twenty statements involved each of the following: Material comfort, sex, harm

¹ These analyses were aided by grants from the Louisiana Graduate Council on Research. The author wishes to thank Kim Ki Suk, George Palmer, and Austin Flint who carried out the statistical analyses. He is also indebted to Donald Campbell, Gerald McCullough, Herbert Rothschild, Arnold Gebel, Perry Davis, Cecil Gibb, and Sam Hicks for their assistance in data collection.

² Heterogeneous samples were selected purposefully for the various analyses upon which the scale development was based. It was assumed that differences in one or more of personal characteristics such as age, sex, education, and regional background would yield increased personality variance. This increase would permit maximum discrimination for analysis and test validation. Moreover, the scale would be developed using samples representing the population for which the scale was being constructed.

avoidance, achievement, affiliation, deference, autonomy, aggression, abasement, rejection, nurturance, superego strength, and irritability. Forty additional statements, many opposed in meaning to each other, covered a wide miscellany of matters.

While it was shown elsewhere (4) that performance on the 13 scales could be accounted for by three common factors, in the present study, it was assumed that one general or second-order factor permeated all the scales as well as the 40 statements that had remained unclassified, the tendency to accept or reject the widely varying statements regardless of their content.

Item Analysis

The proportion of all 300 proverbs accepted by an S was used as a gross measure of the criterion against which a brief acquiescence scale was developed by item analysis. The 200 Ss were ordered according to these criterion scores. The performance of the upper 25 per cent of Ss was then contrasted with the performance of the lower 25 per cent on each of the 300 statements.

Fifty-six items emerged which were accepted by at least 40 per cent more of the upper than the lower criterion groups. These 56 statements, which are listed in Table 1, constitute the Social Acquiescence scale. The number of statements accepted among the 56 constitutes an S's Social Acquiescence score. "?" and "Disagree" responses are ignored.

RESULTS AND DISCUSSION

Reliability

A new sample of 50 West Coast residents and 50 Southern college students was selected with a mean age of 28.2 years ($SD = 2.2$). Forty-nine per cent were male. In this sample, the corrected split-half reliability of the 56-item Social Acquiescence scale was .92. A reliability of .81 was found for a more homogeneous sample of 1491 Louisiana college freshmen, using the Kuder-Richardson formula 21. The 56-item scale correlated .95 with tendency to accept or reject the 300 items.

Acquiescence or Acquiescences?

Consistent with the proposition that acquiescence may appear in a variety of independent forms, a correlation of .00 was found for 38 Louisiana State University night-school students between the Social Acquiescence scale and Ss' tendency to accept, agree, or react positively to 12 nonexistent test items which the experimenter supposedly had in

TABLE 1
THE 56-ITEM SCALE OF SOCIAL ACQUIESCENCE

Success against odds is the greatest of American ideals.
Love is the greatest of Arts.
There is no satisfaction without a companion to share it.
Love of the opposite sex makes the world go round.
They never fail who die in a great cause.
He that has many friends need never fear disaster.
Destroyers of tyranny have contributed the most to mankind.
You only injure yourself when you take notice of despised critics.
The only known cure for fear is faith.
Our chief want in life is somebody who will make us do what we can.
Never trust a flatterer.
He who laughs last laughs longest.
No principle is more noble or holy than that of true obedience.
There is nothing which the body suffers which the soul may not profit by.
One false friend can do more harm than one hundred enemies.
No gift is more precious than good advice.
Obedience is the mother of success.
The victory always remains with those who admire rather than with those who criticize.
The greatest fortunes are for those who leave the common path and blaze a new trail for themselves.
'Tis vain to quarrel with our destiny.
To be happy, always stay within the law.
You should give more than you want to give.
Pity is the touch of God in human hearts.
What we win through authority we lose; what we win through consideration we keep.
A sense of duty is the basis of character.
Next to love sympathy is the most divine passion of the human heart.
Stay away from the proud man who is ashamed to weep.
Sweet is the sleep of men with virtue.
One should feel the failures of his friends as if the failures were his own.
Giving is always better than receiving.
He that loses his conscience has nothing left that is worth keeping.
Virtue is a struggle in which we overcome our weaknesses.
He conquers all who conquers himself.
It is difficult to do excellent work without great strain.
Only a statue's feelings are not easily hurt.
Happiness must be won through great effort.
The restless sleeper blames the couch.
Seeing is believing.
Still water runs deep.
Make yourself honey and the flies will eat you.
The grass is always greener in the other fellow's yard.
Most big cows have little horns.
Every man is blind to his own defects.
Jaws are the only part of the body that like to work.
Those in high places are in greater danger than those in lowly ones.
Life is a struggle from beginning to end.
Wild colts make good horses.
Empty heads go with loud talk.
You can't teach an old dog new tricks.
Count your sheep and the wolf will eat them.
Sleep is loved by everyone.
The feeling of friendship is like that of being comfortably filled with roast beef.
Who does not love the opposite sex remains a fool the whole life long.
Better one safe way than a hundred on which you are not sure.
We like best that which lies beyond our reach.
Amusement is the medicine for worry.

mind. Since both measurements were reliable, they appear to be consistent but independent measures of different forms of acquiescent behavior.

Acquiescence and the Antidemocratic Personality

Gaier and Bass completed an unpublished correlational analysis of the performance of 108 Maryland, 82 Kansas, and 71 Louisiana college students on the Social Acquiescence scale, the ethnocentrism or E scale and the authoritarianism or F scale. For the Maryland, Kansas, and Louisiana samples respectively, the F scale correlated with the Social Acquiescence scale: .49, .48, and .16. The E scale correlated with the Social Acquiescence scale .33, .20, and -.06. The E-F correlation for the three samples varied in the same way: .62, .60, and .49. Judgment is reserved concerning sampling effects on the intercorrelations. It is important to note here, however, that much E-F covariance both within and between samples could be accounted for by the mutual correlation of the two measures with Social Acquiescence. The Kansas sample means were highest while the Maryland sample means were lowest on all three scales.

Relations with Other Measures

Inspection of Table 2 and Table 3 suggests a pattern of relations between social acquiescence and miscellaneous aptitude, personality and behavior measures. One might infer that the person high in social acquiescence is an "outward-oriented," insensitive, non-intellectual, socially uncritical individual; in short, a Babbitt—an unquestioning conformer to social demands placed on him. While in many cases samples are too small to provide statistical significance, the findings will be reported briefly as suggestive for future investigation.

Ability and education. Correlations listed in Table 2 between various ability measures and Social Acquiescence tend to be low and negative. Inspection of the group differences shown in Table 3 suggests a fairly strong negative relation between amount of education and Social Acquiescence. College-trained Ss exhibit less acquiescence than those who have not gone beyond high school.³

³ The same negative relationship with educational level was found for the F scale in several studies. (See Cohn, T. S. and Carsch, H. Administration of the F scale to a sample of Germans. *J. abnorm. soc. Psychol.* 1954, 49, 471.) The fact that the F scale and the Social Acquiescence scale are both negatively associated with education conforms to our hypothesis (3) that the F scale is primarily a measure of acquiescence, not authoritarianism.

TABLE 2

PRODUCT-MOMENT CORRELATIONS BETWEEN THE SOCIAL ACQUIESCENCE SCALE AND MEASURES OF ABILITY, SOCIABILITY, AND SOCIAL SENSITIVITY

Sample	Other Measure	Trait Assessed	Correlation with Social Acquiescence
27 industrial supervisors	Otis Gamma	intelligence	-.15
100 college students	ACE	intelligence	-.12
43 grocery salesmen	Wonderlic	intelligence	.00
100 college students	Grade-point average	academic achievement	-.12
77 factory supervisors	Supervisory aptitude battery	supervisory aptitude	-.23
77 factory supervisors	Years of education	education	-.12
41 night school students	Gordon Personal Profile-A	ascendency	.22
25 college students	GZ Temperament Survey-A	ascendency	.06
41 night school students	Gordon Personal Profile-S	sociability	.34
25 college students	GZ Temperament Survey-S	sociability	.27
77 factory supervisors	Leadership questionnaire	favor initiative	.21
25 college students	Leaderless discussion success	initiative	.32
25 college students	GZ Temperament Survey-P	cooperativeness	.25
23 enlisted Marines	Peer nominations	"likes to help"	.45
23 enlisted Marines	Peer nominations	"thinks well of most"	.42
23 enlisted Marines	Peer nominations	"influences others"	.23
23 enlisted Marines	Peer nominations	"rugged individualist"	-.29
23 enlisted Marines	Peer nominations	"chip on shoulder"	-.24
25 college students	Kerr Empathy Test	empathy	-.14
25 college students	Accuracy on "How Others Rate Me"	empathy	-.34
25 college students	GZ Temperament Survey-T	introversion	-.25

TABLE 3

MEAN DIFFERENCES IN SOCIAL ACQUIESCENCE AMONG SELECTED SAMPLES

Sample	Region	Mean Acquiescence	SD
Adults			
43 Wholesale grocery salesmen	South	38.6	8.51
96 Penitentiary psychopaths	South	34.6	8.99
51 Penitentiary inmates without psychopathy	South	33.8	7.82
77 Industrial supervisors	South	33.7	9.18
18 Wholesale cosmetic salesmen	Midwest	33.6	8.33
34 Marine enlisted men	U. S.	31.9	10.40
49 Public school teachers	South	31.3	9.33
49 Residents	West Coast	31.2	11.24
Adolescents			
78 Rural high school seniors	South	38.3	9.36
145 Rural high school sophomores	South	36.7	8.97
36 Student nurses	New England	35.2	9.06
1,491 College freshman	South	32.6	8.09
100 College sophomores	South	32.1	11.60
100 College sophomores	Midwest	29.6	11.00
45 College seniors and graduate students	South	26.9	10.49

Personality. Correlations listed in Table 2 suggest that the socially acquiescent individual tends to be more sociable and yet more insensitive socially.

Population differences. The belief that the

acquiescence scale is a measure of "Babbitttry" is supported by a further examination of some of the differences among the groups listed in Table 3. Southern salesmen, much more likely to have Fromm's "marketable personality," show more Social Acquiescence than any other Southern group. Similarly, the small Northern group of cosmetics salesmen appear to be fairly high in Social Acquiescence for a non-Southern adult sample. Prisoners tend to be less acquiescent than nonprisoners matched for educational level (13). The greater tendency to acquiesce (about one third of a standard deviation) of Southern compared with Midwestern sophomore college students is consistent with their differences in background. In the South, there remains a more ready acceptance of the traditional and conventional manifested in the greater religiosity among Southern students as a whole, "official optimism" about social issues, and lack of skepticism.

One last analysis that supports the "Babbitttry" hypothesis involved examination of the scores of 45 advanced college students who were asked to take the Social Acquiescence test twice, once imagining themselves applying for a job where they wish to appear as de-

sirable as possible, and once answering the questions as honestly as possible. The mean "faked" score of 30.3 was significantly greater at the .01 level than the mean "honest" score of 26.9, supporting the inference that high Social Acquiescence is regarded as desirable just as sociability, ascendancy, emotional stability, and other such traits are usually thought by the layman to be related positively to career success.

Research Significance

The evidence presented on Social Acquiescence suggests the need to reinterpret much published social psychometric data. It is probable that Social Acquiescence accounts for much, if not all, of the covariance found between a wide variety of measures ranging from favoring birth control to belief in the efficacy of Yoga or conversely, disapproving desegregation to rejecting the use of American military force in Outer Mongolia.

There are at least four ways of eliminating or reducing the constant error due to acquiescent tendencies. First, new social psychological measurements can be constructed to balance carefully affirmative items with their opposites. Second, the reading ease of items can be reduced. Third, more intelligent and critical subjects can be employed. Fourth, acquiescence can be measured and then controlled by means of partial correlation or by sampling procedures.

SUMMARY

A 56-item scale of social acquiescence has been developed by internal consistency item analysis. It has a reliability above .90 and correlates .95 with the tendency to accept or reject 300 proverbs, aphorisms, and similar statements. According to population differences and correlations between the scale and various personality and intelligence measures, one who earns a high Social Acquiescence score resembles Sinclair Lewis' "Babbitt"—an unquestioning conformer to social demands. The Babbitt syndrome still must be regarded as a

tentative sketch of the socially acquiescent S since many of the correlations reported were based on small samples and lacked statistical significance.

The demonstrated existence of consistent individual differences in tendency to acquiesce to generalizations about behavior has methodological importance. Previous research results based on the correlation between attitude scales (e.g., E and F scales) in which all statements are scored in the same direction, require reinterpretation.

REFERENCES

1. ANCONA, L. Indagine sulla natura psichica del "Response Set." La "motivazione al successo." *Arch. psicol. neurol. Psichiat.* 1954, **15**, 23-74.
2. BARNES, E. H. The relationship of biased test responses to psychopathology. Unpublished doctor's dissertation. Northwestern Univer., 1954.
3. BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.*, 1955, **51**, 616-623.
4. BASS, B. M. Development of a structured disguised personality inventory. *J. appl. Psychol.*, in press.
5. BERG, I. A. Response bias and personality: the deviation hypothesis. *J. Psychol.*, 1955, **40**, 61-72.
6. BERG, I. A., & COLLIER, J. S. Personality and group differences in extreme response sets. *Educ. psychol. Measmt*, 1953, **13**, 164-169.
7. BERG, I., & RAPAPORT, G. M. Response bias in an unstructured questionnaire. *J. Psychol.*, 1954, **38**, 475-481.
8. COHN, T. S. Factors related to scores on the F (Pre-disposition to Fascism) Scale. Unpublished doctor's dissertation. Univer. of Michigan, 1953.
9. CRONBACH, L. Further evidence on response sets and test designs. *Educ. psychol. Measmt*, 1950, **10**, 3-31.
10. FLEISHMAN, E. The measurement of leadership attitudes in industry. *J. appl. Psychol.*, 1953, **37**, 153-158.
11. GAIER, E. L., LEE, M. E., & MCQUITTY, L. L. Response patterns in a test of logical inference. *Educ. psychol. Measmt*, 1953, **13**, 550-567.
12. LORGE, I. Gen-like: Halo or reality? *Psychol. Bull.*, 1937, **34**, 545-546.
13. PALMER, G. Discrimination of psychopaths, normal prisoners and non-prisoners using a disguised objective personality test. Unpublished master's thesis, Louisiana State Univer., 1956.
14. SINGER, W. B., & YOUNG, P. T. Studies in affective reaction: III. The specificity of affective reactions. *J. gen. Psychol.*, 1941, **24**, 327-341.

Received December 15, 1955.

PRIDE IN GROUP PERFORMANCE AND GROUP-TASK MOTIVATION¹

LEONARD BERKOWITZ

University of Wisconsin

AND BERNARD I. LEVY

Crew Research Laboratory, Air Force Personnel and Training Research Center

INVESTIGATIONS conducted by the Survey Research Center (SRC) have shown that the attitude constellation variously labeled "pride in work group" (e.g., 4) or "pride in group performance" (7) tends to be related to group productivity. In discussing this relationship, investigators (e.g., 5) have suggested that work-group members develop pride in their group's ability to perform the designated task as a result of perceptions of past performance effectiveness, and that the resulting attitudes then contribute to future performance effectiveness.

The principal criterion employed in the SRC studies has been workgroup "productivity." It is plausible to assume that task motivation is among the major determinants of this variable. On this assumption, two aspects of the hypothesized relationship between pride in group performance and task motivation can be distinguished. In the first place, since attitudes toward the group's performance are explicitly at issue, the relevant favorable attitudes held by any individual member supposedly result from his perceptions of the group's effectiveness rather than of his own effectiveness alone. A measure of pride in group performance therefore should be affected more strongly by evaluations of the performance of the group as a whole than by evaluations of the separate group members, particularly when the individual members have no cues as to how well each of the others in their group is doing.

The second and major consideration in the hypothesized relationship is that pride in

group performance is assumed to be associated with high motivation to perform the task assigned to one's group, and precedes later high productivity. To our knowledge there is little direct evidence on this last point. A study reported by the Crew Research Laboratory (6) has demonstrated that aircrews in training that have high "pride in crew" (which is conceptually similar to pride in group performance) tend to be rated highly by their superiors in later combat assignments. However, since the superiors' ratings may depend on factors other than task motivation (such as interpersonal harmony among the crew members) it is at least conceivable that this finding may not be generalizable to the relationship between group attitudes and task motivation.

The general purpose of the present experiment is to test the relationship between pride in group performance and task motivation under carefully controlled conditions. Subjects' (Ss') "pride" in their group's performance was manipulated by means of arbitrary performance evaluations. The evaluations varied in their nature (whether they were favorable or unfavorable) and in their target (whether they were directed to the group as a whole or to each of the individual members). An analysis was conducted to determine the effects of the nature and target of the performance evaluations upon pride in group performance. The different conditions then were compared in terms of an index of task motivation. The hypothesis, derived from the SRC studies, was that groups receiving favorable evaluations of the group as a whole would score higher on this motivational index than groups receiving unfavorable evaluations of the entire group or groups in which the members receive evaluations of solely their own performance. The effect of subsequent task performance upon attitudes toward the group was finally investigated.

¹ This study was performed under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas in support of Project 7713. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views or have the endorsement of the U. S. Air Force or U. S. Government.

METHOD

Subjects. The present sample consists of 25 groups, each consisting of three airmen drawn from a population of basic trainees. The airmen varied markedly in intelligence classification but included substantial representation from the highest intelligence groups.

Group task. The group task was employed solely as a convenient setting within which the effects of performance evaluations could be studied. The assigned task was based upon Lanzetta's modification of a task first employed by Schutz,² involving a miniature 3-man Air Defense warning center. Three discrete target areas were located near the center of a metal-covered board with a grid identified by coordinates. Model planes with small attached magnets represented "enemy" and "interceptor" air forces. The following types of information were supplied to each group: (a) A continuous tape-recorded message, divided into approximate 4-minute intervals, reported the coordinates of all planes (other than friendly interceptors): 15 unidentified aircraft, 9 enemy bombers and 6 friendly planes. During each interval the group had the task of translating the aircraft positions reported into moves on the interceptor board, and deciding on the deployment of interceptors. (b) A second input consisted of flight plans for the 6 friendly planes posted on the "Friendly Aircraft Status Board" near the metal-covered board. The general requirement was to defend the three target areas against an attack by enemy bombers. To do so necessitated correct identification of friendly and enemy planes, and strategic deployment of the available interceptor force for counterattack.

Procedure. All Ss first were told that the purpose of the experiment was to study the way in which people work together in important Air Force groups such as Air Defense teams. Detailed instructions were given about the sequence of activities which included a brief practice session, an initial trial, a 15-minute "break" period, and finally a second trial. The Ss in the experimental groups were advised that psychologists and experts would observe their performance from behind a one-way mirror and evaluate them by comparing their performance with that of over 200 groups of basic trainees like themselves who had worked previously on the task. They were informed that they would receive psychological evaluations at the end of the fourth and eighth intervals of the first trial. Nothing was said to the control groups about their being evaluated. An "objective" scoring scheme was then explained to both experimental and control Ss: they would receive 20 points for every enemy bomber downed, lose 50 points for every target hit, and lose 20 points for every friendly plane downed and fighter lost.

After a brief practice trial of two intervals, the Ss were given the first trial, which was stopped before the end of the problem, at the completion of the eighth interval. The Ss then filled out a brief attitude and sociometric questionnaire, after which they were told

they could take a 15-minute "break" but that they would have to remain in the room.

The *E* then went to the adjoining room where he observed the Ss' behavior during this rest period through a one-way mirror. Magazines were displayed on a table at one end of the room for the Ss to read if they desired, although nothing was said to them about this. The *E* categorized their behavior and discussions.

At the end of 15 minutes the *E* returned to the experimental room and started the Ss upon the second trial. Upon completion of the problem the Ss were told how many enemy bombers had been downed and how many targets had been hit. The questionnaire then was readministered and the Ss dismissed.

Experimental manipulations. There are five conditions of five groups each. Four of the conditions represent a factorial design with the following sources of variance: (a) nature of evaluations; (b) target of evaluations; (c) interaction of nature and target; and (d) residual. Half of the experimental groups received favorable evaluations of their performance (+) during the first trial while the other half of the sample received unfavorable evaluations (-) during this trial. Similarly, half of the experimental groups received evaluations of the group as a whole (G); the individual members were not singled out and each member saw the group rating. The Ss in the remaining evaluated groups (I) received ostensible evaluations of their own individual performance. The Ss in the I conditions were instructed not to communicate with each other concerning their ratings. Evaluation of both groups and individuals was in terms of percentile scores on ratings of seven attributes such as: (a) understanding of problem requirements; (b) utilization of resources; (c) coordination; and (d) timing and pacing of work. It was believed that these attributes would seem meaningful to the Ss but yet would be nebulous enough so that the Ss would not have any objective basis for questioning the ratings. The ratings given the praised groups averaged 85 at the fourth interval report and 90 at the eighth interval. For the criticized groups the ratings averaged 35 and 30 respectively.

The fifth condition was a control (C), included in order to examine the effects due to the evaluations per se, regardless of their nature. Groups in this condition were not told that they were to be observed and evaluated and were not given any ratings.

Observations. The observations made by the *E* during the 15-minute "break" period consisted of 15-second time samples taken each minute. Recording was in terms of five categories: "Keeps to self" (S); "Task-oriented discussion" (T-O); "Non-task-oriented discussion" (N-T-O); "Critical remarks" (C); and "Remarks concerning the present experiment" (i.e., what the purpose was, what the experimenter would do next, etc.) (E). No critical remarks were ever heard. Since very few remarks were coded in Category E, this category was not used in the analysis.

The observers coded any *S* reading a magazine or otherwise not engaged in conversation during the time sample as "keeping to himself." A remark made at any time during the 15-second period was coded in one of the three verbal categories according to its predominant function. For the present analyses, the total number of units coded in any one category over all

² We are indebted to Dr. John Lanzetta for the construction of the Air Defense task and for suggestions for the modification of the task for the present experiment.

15 time samples are employed as group scores. The number of units coded as task-oriented discussion (T-O) is taken as an index of the group's task motivation. This measure, the major dependent variable, is an index of the group members' concern with their assigned task at a time when they are under little overt pressure to be so concerned.

An estimate of the reliability of observations was computed using data from five randomly selected groups observed by both Es, involving an analysis of variance design in which the classifications were categories, observers, and groups. The continuous variable was the group score in each category. The Hoyt (1) reliability estimate from this analysis was .99.

Attitude and sociometric measures. The attitude questionnaire administered immediately after each trial contained one item designed to test the success of the experimental manipulations, and nine items to assess pride in group performance. Five of the nine items were selected for the pride-in-group scale after applying the iterative scaling technique described by Stice and Knoell (6) to the post-first-trial responses. All analyses were based upon group scores on these five items. In combining the items, scores on each item first were converted into *T* scores. The group-scale score was the sum for all members over all items. A constant of 200 was subtracted from each group-scale score to facilitate the analysis. Two of the items in the scale are:

If you were going to go through this exercise again, would you expect to do better than: (Check one)

- 20 per cent of the other people going through it
- 40 per cent of the other people going through it
- 60 per cent of the other people going through it
- 80 per cent of the other people going through it
- All others.

How well do you think each of your present teammates has worked on the problem you have just completed? (Poor job, 1, through Excellent job, 5.)

Objective performance measures. The objective measures employed here were number of "enemy bombers" downed before hitting a target, and the number of "friendly aircraft" accidentally shot down. We believe that these performance measures are dependent largely upon intellectual and problem-solving processes rather than group-task motivation. The performance data were collected, not as criteria or for validation of the motivation index, but in order to study the effects of quality of performance during the second trial upon subsequent changes in pride in group performance.

RESULTS

Success of the experimental manipulations. The questionnaire item designed to test the success of the manipulation reads as follows:

What is your impression of your group's effectiveness as an Air Defense team as compared with the effectiveness of other 3-man teams composed of basic

- trainees going through this exercise? (Check one):
- Better than 20 per cent of the other teams
 - Better than 40 per cent of the other teams
 - Better than 60 per cent of the other teams
 - Better than 80 per cent of the other teams
 - Better than all others.

The weight given to each alternative was the percentage of other teams thought by each *S* to have been less effective than his team. The score for each group was obtained by summing the scores for each of the three members. The mean group score immediately after the first trial in each of the five conditions is shown in Table 1.

Analysis of variance of the data in the four experimental conditions suggests the following: there is no significant difference between the perceived effectiveness scores of the *G* groups as compared with the *I* groups ($F = 0.2$). However, the favorably evaluated groups, as would be expected, have much higher scores on this measure than the unfavorably evaluated groups ($F = 1,071.6$). There is also a significant interaction ($F = 6.69$; $p = .05$), which may be a function of the amount of information provided to the *Ss* concerning the performance effectiveness of others in the group.

In comparison with the control groups' scores, the groups in the two favorably

TABLE 1
MEAN PERCEIVED GROUP-EFFECTIVENESS SCORES IN EACH OF THE FIVE CONDITIONS

Target	Nature of Evaluations	
	Favorable (+)	Unfavorable (-)
Entire group (G)	220	100
Individual members (I)	192	120
Control condition:	144	

TABLE 2
MEAN "PRIDE-IN-GROUP" ATTITUDE SCORE AFTER THE FIRST TRIAL IN EACH OF THE FIVE CONDITIONS

Target	Nature of Evaluations	
	Favorable (+)	Unfavorable (-)
Entire group (G)	58.93	45.82
Individual members (I)	54.69	43.60
Control condition:	45.39	

evaluated conditions had perceived effectiveness scores significantly greater than the control condition ($t = 5.07$, $p = .001$, for the G+ groups; and $t = 3.20$, $p = .01$ for the I+ groups), while the G- groups had perceived effectiveness scores significantly lower than the control groups ($t = 2.93$, $p = .05$).

Pride in group after the first trial. Table 2 presents the mean pride-in-group score in each of the five conditions for the responses made immediately after the first trial. Analysis of variance of the data for the experimental conditions shows that the favorably evaluated groups tend to have significantly higher attitude scores than the unfavorably evaluated groups ($F = 45.96$, $p = .001$). The main effect for target, G or I, while not significant ($F = 3.29$, $p = .10$), indicates a trend in the direction of our hypothesis.

The t test of the difference between the G+ condition and the pooled groups in the three other conditions (using a one-tailed test)³ is highly significant ($t = 5.34$; $p = .001$). This condition also has a significantly higher score than the next highest condition, I+ ($t = 2.08$; $p = .03$). Thus, the pride in group performance measure, as predicted, is more strongly affected by the perceived merit of the total groups's performance than by each member's perception of his own individual performance. Lastly, groups in both favorably evaluated conditions tend to have somewhat higher pride-in-group scores than the control groups, while there are no significant differences between the control and unfavorably evaluated groups.

Behavior during the "break" period. The observers' frequency tallies were found to have a skewed distribution and, therefore, were normalized by the square-root transformation. The mean transformed scores for each of three categories are given in Table 3. It can be seen from this table that the ordering of conditions in terms of pride in group performance is not the same as their ordering on this measure of task motivation, although the G+ groups are high on both variables.

Analyses of variance of the data in the four experimental conditions indicate that the target of the evaluations has a significant effect upon the T-O and S categories of

TABLE 3

TOTAL NUMBER OF BEHAVIORAL UNITS IN EACH OF THE THREE CATEGORIES DURING THE "BREAK" PERIOD
(Square-root transformation)

Target	Nature of Evaluations					
	Favorable (+)			Unfavorable (-)		
	T-O*	N-T-O†	S‡	T-O	N-T-O	S
Entire group (G)	23.4	13.7	12.6	18.6	16.7	15.2
Individual members (I)	13.9	18.3	19.2	10.9	14.9	23.1
	T-O		N-T-O		S	
Control condition:	14.2		12.1		24.6	

* Task-oriented discussion

† Non-task-oriented discussion

‡ Keeping to himself

behavior ($F = 4.52$, $p = .05$ and $F = 4.89$, $p = .05$ respectively) but not on the N-T-O category. Task-oriented discussion is more frequent when the evaluations—favorable or unfavorable—are given to the group as a whole, and more Ss keep to themselves when evaluations are given to individual group members. The level of motivation on the present interdependent group task apparently is influenced by whether evaluation is of the group as a whole or of each individual's own performance. Furthermore, the high frequency of Ss keeping to themselves in the I conditions suggests that individual evaluation tends to divide the group.

We had hypothesized that the groups in the G+ condition would have the highest scores of all conditions on the task-motivation index. This hypothesis was tested by t tests of the difference between this condition's means on the observational categories and the means obtained by pooling the groups in the other three conditions, basing the error terms upon the residual mean squares. The resulting t 's are significant for categories T-O and S. They are 1.92 ($p = .04$) and 1.75 ($p = .05$) respectively. Thus, the G+ groups tend to have more task-oriented discussions and fewer instances of members keeping to themselves than the 15 groups in the three other conditions.

The t tests were also made to determine if the evaluated experimental conditions differed from the nonevaluated control conditions in the frequencies of behavior in the three categories. There were no significant differ-

³ All t tests specified by the a priori hypotheses employ one side of the distribution.

ences between any of the evaluated conditions and the control condition for the T-O category. A significant tendency is found, however, for the G+ condition to result in fewer instances of members keeping to themselves (S) than the control condition, a finding consistent with the interpretation that favorable evaluation of the group as a whole has served to unite the members of the group.

In summary, the tendency for the G+ group to be different from the groups in the other conditions on both pride in group performance and degree of concern with the task suggests that there may be two sets of factors involved in the present attitudes related to motivation to perform the assigned group task. The high pride that is associated with high group-task motivation apparently results from perceptions of (a) effective performance, (b) by the group as a whole.

Objective performance measures for the second trial. There were no statistically significant differences in group performance during the second trial. As explained earlier, performance on this Air Defense task was thought to be relatively independent of task motivation.

Pride in group after the second trial. The ordering of conditions in terms of the post-second-trial attitude scores is not similar to the post-first-trial ordering, and there are no significant differences among the five conditions. There are also no significant effects for the four experimental treatments or their interaction.

A comparison of the scores obtained in the different conditions before and after the second trial indicates that the unfavorably evaluated groups had increased in their "pride" scores, while the favorably evaluated groups had generally shown a decrease in these scores. An attempt was made to determine if the change in the attitude scores was related to the initial attitude level, later performance, or the interaction of these two variables.

The distribution of scores on each of the two variables, post-first-trial pride in group and second-trial performance effectiveness, was first dichotomized. (The performance measure was obtained employing the weighting scheme described to the Ss). The 25 groups then were divided among the four conditions

TABLE 4

MEAN CHANGE IN PRIDE IN GROUP IN EACH OF THE FOUR COMBINATIONS OF POST-FIRST-TRIAL PRIDE IN GROUP AND SECOND TRIAL PERFORMANCE EFFECTIVENESS

Performance Effectiveness	Post-First-Trial "Pride in Group"	
	High	Low
High		
<i>N</i>	6	5
Mean change	-4.1	5.3
Low		
<i>N</i>	6	5
Mean change	-5.6	4.6

of the resulting factorial design, omitting three groups selected at random from the Low First-Trial Attitude-High Performance condition in order to obtain proportional cell frequencies. The mean change in attitudes from first to second trial in each of the four conditions is given in Table 4.

The data, arranged in the design shown in Table 4, were subjected to an analysis of variance. Two of the three effects are statistically significant. 1. The groups with initially high attitudes tend to develop less favorable attitudes while the groups with lower attitude scores tend to develop more favorable attitudes ($F = 9.91, p = .01$). This regression effect suggests that the pride in group performance brought about by the present experimental manipulations is relatively unstable. 2. The interaction of initial attitude score and performance also is significant ($F = 9.77, p = .01$). In other words, the groups with initially high pride in group but relatively poor later performance tend to decrease the most in their attitude scores, and the groups with initially low pride in group but relatively good later performance show the greatest increase in attitude scores.

DISCUSSION

It has been assumed that the amount of task-oriented discussion within a group during the work groups' "break" period provides an index of group-task motivation. The groups were under little overt pressure to be concerned with their task during this rest period and the presence of magazines in the room implied that they were free to read or engage in other non-task-related activities. Furthermore, since there were no between-condition

differences in objective performance prior to the "break" period, differences in the amount of task-oriented discussions can safely be attributed to the performance evaluations given to each condition. The obtained results, then, support the implications inherent in the SRC hypothesis. The pride in group performance associated with high group-task motivation results from group members' perceptions of *group* effectiveness rather than from each member's perception of his own effectiveness alone.

What processes produce the relatively high group-task motivation under conditions of favorable group evaluation? Katz and Kahn (5) suggest that the high group-task motivation associated with pride in group performance is a direct result of increased group attractiveness. However, experimental evidence indicates that a high level of group attractiveness can be as readily associated with low group-task motivation as with high motivation depending upon such other factors as group standards for low or high productivity (1, 8). In those groups having a standard for high production, groups highly attractive to their members should be more highly task-motivated than less attractive groups. The attainment of the group goal of high productivity by the former groups, furthermore, may well result in relatively high pride in group performance. But a problem remains. Why do some groups develop standards for high production and not others?

A better explanation, we believe, can be found by considering the two aspects of the evaluations, their nature and their target. Favorable evaluations should reward the group members, but differences in the target of the evaluations result in different things being rewarded. If the *individual* members are favorably evaluated, they receive rewards for their *individual* performances. If the group is favorably evaluated each *S* receives a reward only by virtue of his being a member of the group. In other words, the favorable evaluations of the entire group serve to create the perception that the members are interdependent in attaining rewards; each *S*'s reward attainment is dependent upon the others in his group. Research findings indicate that this type of perception can produce relatively high group-task motivation. Deutsch

(2) experimentally established "cooperative" and "competitive" group situations, differing in the extent to which experimental conditions favored the mutual perception of "promotive interdependence" among the members. Members of the "cooperative" groups felt stronger obligation to each other and were more highly motivated to work hard at the group task than did the "competitive" group members. The relationship between pride in group performance and task motivation may result, thus, from a process in which group members high in pride perceive that each member's attainment of the rewards associated with effective performance is dependent upon the others in the group.

Two sets of evidence obtained by the present investigators support the contention that the target of the performance evaluations can affect the members' perceptions of interdependence. First, the "individual evaluations" condition in the present experiment has a greater incidence of members keeping to themselves in the "break" period than the condition receiving "group evaluations," while the *G+* groups had the smallest amount of this type of behavior. Although alternative explanations can be found, it is plausible to assume that group members perceiving a high degree of member interdependence for goal attainment would be relatively unlikely to keep apart from the others in the group. Second, a yet unpublished study compares the condition in which members are aware that everyone in the group has received favorable individual evaluations with the condition in which favorable evaluations are given to the group as a unit. As was argued above, members of groups in the "individual evaluation" condition should not perceive the attainment of their rewards as being greatly dependent upon the others in the group even though they may be aware that everyone has done well. The results are in accord with this hypothesis. The former groups had a smaller incidence of task-oriented discussions and a greater incidence of members keeping to themselves than the groups favorably evaluated as a unit.

SUMMARY

Previous studies have found that work groups having high pride in group performance

tend to be relatively highly productive. It generally is assumed that the attitudes have contributed to the effective performance. The present experiment was designed to test implications of this hypothesized relationship, namely, that pride in group results from group members' perceptions of (a) effective task performance of (b) the group as a whole (rather than of their own separate performances) and is directly associated with high motivation toward the assigned group task.

Groups in the present study were assigned to work on a simulated Air Defense task. They received an initial trial, on the basis of which they supposedly were evaluated, a "break" period, and then a second trial. Twenty-five 3-man groups were distributed among five conditions: two conditions in which group performance on the initial trial was favorably evaluated, and two conditions in which this performance was unfavorably evaluated. One of the favorably evaluated and one of the unfavorably evaluated conditions received evaluations of the group as a whole while the other conditions received evaluations of the individual members. In each case the evaluations were provided through previously constructed ratings. A final, nonevaluated condition served as a control.

The amount of task-oriented discussion during the "break" period was used as the measure of task motivation. Groups receiving favorable evaluations of the group as a whole were found, as predicted, to have higher pride in group performance and greater concern with the assigned group task than groups having either unfavorable group evaluations or evaluations, whether favorable or unfavorable, of the separate group members.

Evidence is cited supporting the hypothesis that the high group-task motivation related to high pride in group performance results from a perception of interdependence among the group members with respect to the attainment of reward. Conditions in which group members presumably believe they can be rewarded for effective performance somewhat independently

of the others in the group result in a relatively high incidence of members keeping to themselves during the "break" period and in relatively low group-task motivation.

The attitude responses made after performance on the second trial yielded the following results: The level of a group's pride in group established at the termination of the first trial was relatively unstable. There was a significant tendency for the groups with highly favorable attitudes to develop less favorable attitudes, and for the groups with low scores to increase in their "pride." The level of pride in group changed if the initial level was at variance with the quality of the group's performance, an effect that may be dependent, in part at least, upon the relative instability of the initial attitude level.

REFERENCES

1. BERKOWITZ, L. Group standards, cohesiveness and productivity. *Hum. Relat.*, 1954, **7**, 509-519.
2. DEUTSCH, M. The effects of cooperation and competition upon group process. *Hum. Relat.*, 1949, **2**, 199-231.
3. HOYT, C. J. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.
4. KATZ, D., MACCOBY, N., & MORSE, NANCY. *Productivity, supervision and morale in an office situation*. Ann Arbor: Survey Research Center, Univer. of Michigan, 1950.
5. KATZ, D., & KAHN, R. L. Some recent findings in human-relations research in industry. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 650-665.
6. KNOELL, DOROTHY M. Relationships between attitudes of bomber crews in training and their attitudes and performance in combat. *USAF, Personnel Train. Res. Cent., Res. Bull.*, in press.
7. MORSE, NANCY. *Satisfactions in the white collar job*. Ann Arbor: Survey Research Center, Univer. of Michigan, 1953.
8. SCHACHTER, S., ELLERTSON, N., MCBRIDE, DOROTHY & GREGORY, DORIS. An experimental study of cohesiveness and productivity. *Hum. Relat.* 1951, **4**, 229-238.
9. STICE, G. F., & KNOELL, DOROTHY M. A simple mean-difference technique for obtaining scales. *USAF Hum. Resour. Res. Cent., Res. Bull.*, September 1953, No. 53-36.

Received December 19, 1955.

EFFECTS OF WORK-GROUP STRUCTURE AND CERTAIN TASK VARIABLES ON GROUP PERFORMANCE¹

JOHN T. LANZETTA

Fels Group Dynamics Center, University of Delaware

AND THORNTON B. ROBY

Institute for Applied Experimental Psychology, Tufts University

ONE feature characterizing many complex man-machine systems is the fact that relevant information regarding the condition of the system, which is necessary for various control actions, may not be available at a single vantage point. Even if readily accessible, the volume of information to be processed is generally so large as to make it virtually impossible for a single individual to monitor more than a small fraction of the relevant input displays. Since decisions and control actions typically require information about the status of many system variables, such systems are of necessity manned by teams in which each individual serves both as an operator, or control agent, and as an observer, or information source for himself and other control agents. Where face-to-face interaction is impossible this requires a communication network so that each operator's information can be supplemented by information relayed by other group members.

It is a plausible hypothesis that in systems of this sort performance will be affected by the volume of information to be relayed, the extent to which relevant information is dispersed throughout the group, and the nature of the communication "net" supplied for information transmission. The latter variable has received considerable research attention in recent years (e.g., 9) but the effects on group performance of variation in volume and dispersion of information have received little systematic exploration. The present study is one in a series directed toward investigation of these particular variables.²

¹This research was performed in support of ARDC Project 7713 under the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government.

²A discussion of the framework on which the studies are based is presented in an unpublished paper; J. T.

A previous study (7) examined the performance of three-man groups operating under four imposed communication structures. Each of the structures was symmetrical in that all three group members had identical task requirements, and all entailed the processing of the same amount of information and the execution of equivalent control adjustments. The structures differed only in the proportion of information which had to be transmitted and in the extent to which relevant information was dispersed over group members. The principal finding of this study was that performance could be related in a meaningful way to the volume of information which had to be transmitted and to the nature of the sources of the information on which control decisions were based. If relevant information was directly available to a response "agent" in the form of a visual display, or if relayed information could be obtained from a single outside source, performance on controls was markedly better than when information had to be obtained from several external sources.

The present study is intended to extend this investigation of structure while at the same time exploring the importance of certain task variables. Considerable research has been directed to the determination of the information-handling capacity of individuals (e.g., 5), but little parallel research has been done on teams. How does team performance vary as a function of information processing demands? Does the information-handling capacity of teams depend on the requirements for internal transmission of information imposed by a communication structure? Both of these questions are of interest in the present investigation.

Lanzetta, and T. B. Roby, "A framework for the study of work-group structure and task parameters," Laboratory Note, Crew Research Laboratory, LN-55-9, June 1955, and a brief discussion of the general theoretical model underlying the approach is presented in a forthcoming publication (8).

The two structures employed in the study were selected from the four used in the previous experiment (7). They represent the conditions which differed most markedly in the extent to which they required information transmission. The two task variables involved operations on the input to the group. A differential and varying characteristic of many group tasks is the amount and predictability of the inputs which must be handled. Both the rate of change of the input and the predictability of the locus of a change were varied in the present study.

METHOD

Apparatus and Task

Since the apparatus and general features of the task have been previously described in detail (7), they will be summarized briefly here. Each of the three Ss belonging to a group was seated in a separate booth which prevented visual and direct auditory communication between group members. Each booth contained a standard aircraft interphone circuit with a throat mike and headset, and all communication between group members was by means of this circuit. By depressing a hand switch, any S could communicate simultaneously with both other Ss.

In each booth there were two switches, each having an "off" setting and three "on" positions. Near each switch was a printed table of operating instructions stating the correct switch settings for various combinations of instrument readings. A LaBelle automatic slide projector behind each S was used to project a picture of two simulated aircraft instruments on the front wall of the booth. The projectors permitted stacking of about 30 slides which were tripped into the projection chamber by a remote-control switch operated by E. The pairs of simulated instruments presented to the Ss were different in each booth.

The correct setting for each control could be determined only by reference to joint readings on two instruments. Conversely, each instrument reading was

relevant to the correct settings for two controls (see Table 1). The general task set the group was to process the information presented (instrument readings), relaying the necessary information to the proper booths, and to execute control actions based on relayed or directly available instrument readings. Each of the three "on" positions for a given switch was connected to one pen on a 20 pen Esterline-Angus recorder so a complete time record of the six control settings made by the group could be obtained.

Experimental Conditions

Structure. Two structure conditions were employed. In the Low Autonomy Structure (LA) each S had two controls for which none of the four necessary instrument readings was directly available. In the High Autonomy Structure (HA) each S had two controls for which all but one of the requisite instrument readings were available in his booth. Thus, LA required each response agent to obtain four units of information from external sources, while HA required each S to obtain one unit from external sources. In addition, in the LA Structure the information required by each member was distributed among both other group members while in the HA Structure the required information was available from a single external source. The interdependencies between group members are represented in Table 2 by a matrix schema.

With the apparatus employed it was most convenient to fix the location of instruments within a booth for all conditions. Thus, slides showing Compass Heading and Air Temperature were presented in Booth 1; Generator Voltage and Fuel Pressure in Booth 2; and Air Speed and Rate of Climb in Booth 3. It also seemed advisable to maintain the same fixed dependencies between control settings and instrument readings for all conditions as indicated in Table 1. Hence structure was manipulated by changing the location of the controls and involved only the moving of control name plates and instruction cards from one booth to another. It can be readily seen from examination of Table 1 that with a fixed allocation of instruments changing the location of controls varies the required degree of communication.

Task variables. Three input parameters were ma-

TABLE 1
INFORMATION BASES OF EACH OF THE SIX CONTROLS

Control	Relevant Information					
	O ₁ Compass heading	O ₂ Air temperature	O ₃ Generator voltage	O ₄ Fuel pressure	O ₅ Air speed	O ₆ Rate of climb
C ₁ . Power setting						
C ₂ . Steering mechanism	x	x				
C ₃ . Landing gear	x					
C ₄ . Control switch				x		
C ₅ . Selector knob			x	x		
C ₆ . Reset lever			x			x
		x			x	x
					x	

Note.—An "x" in a given row and column indicates that a reading on the instrument at the column heading is relevant to decision for the control in the row.

TABLE 2
INFORMATION TRANSMISSION REQUIREMENTS IN THE
TWO EXPERIMENTAL STRUCTURES*

	Source of Information			
	Group Member	A	B	C
High Autonomy Structure (HA)				
User of information	A	3	1	0
	B	0	3	1
	C	1	0	3
Low Autonomy Structure (LA)				
User of information	A	0	1	3
	B	3	0	1
	C	1	3	0

* The entries indicate the items of information (instrument readings) held by the group member indicated in the column and required by the individual indicated in the row. Entries in the diagonal indicate items directly available. For correct control adjustments each of these items of information had to be transmitted for every sequence of three slide changes.

nipulated. The input to the group consisted of slide changes which altered the instrument readings within a booth. Nine different slides (two different instrument readings per slide) were presented in each booth by means of the LaBelle projector. The order in which slide changes were made in the three booths was staggered so that the total information display for the group changed 27 times, e.g., a slide was changed in Booth 1 then, a fixed interval later, in Booth 3, etc. The interval between successive slide changes, irrespective of booths, was one of the variables manipulated. Changes were made at either 15-second or 10-second intervals.

The second independent task variable was the predictability of the change sequence. The order in which slide changes were presented in the booths was either "random" (subject to the restriction that each booth received a slide change before the next sequence was initiated), or "predictable." In the former case the order of slide changes for the three booths was determined from a table of random numbers, in the latter case the order of change was always Booth 1, Booth 2, Booth 3.

The third task variable was of peripheral interest and was included primarily as a methodological check. The 30 slides for each booth were arranged in two different sequences resulting in different required patterns of control changes.

Subjects and design. The Ss were Air Force basic trainees undergoing training at Lackland Air Force Base. Six Ss were available for each experimental day; the six were divided fortuitously into two groups of three Ss each. The experimental conditions were randomly ordered prior to arrival of Ss, and groups were assigned to a condition in the order in which they were received. Pertinent demographic characteristics of the Ss were not available. However, the age range of this population is generally from 18-22 years and the range

of education and intelligence is larger than that in most college samples.

The three task variables were arranged in a $2 \times 2 \times 2$ factorial design with two groups assigned to each cell. Each of the two groups performed under both structure conditions, the order of presentation of the structure conditions being counterbalanced for each cell.

Administration

The Ss were brought into the experimental room, seated in the booths fortuitously, and given a brief introductory talk. They were informed that the experiment would not affect their Air Force status and that the experimental task required them to operate much like members of a bomber crew. It was explained that—as in a bomber—they would make certain control adjustments on the basis of the best available information but that *all* the information usually would not be available to any one of them, so they would need to obtain information from other crew members. After they had been shown how to adjust and use the interphone equipment, they were given three practice slides. No standard practice period was adopted, but the experimental sequences were not begun until the experimenter felt that all group members understood the task.

The information input series described was repeated three times for each group under both conditions. Between repetitions of the series (within conditions) a blank slide appeared, and all Ss reset their controls to "off." At the conclusion of the three periods of the first session, the men were permitted to take a 10- to 15-minute break while the control labels were changed. Thus the total running time for each group was about 20 minutes for the introductory talk; 21 minutes for each of two experimental sessions; and 10 minutes for changing apparatus.

RESULTS AND DISCUSSION

The scores used in the present analysis are the number of errors within three periods, each period consisting of 27 slide changes. Each time a slide was changed in any of the three booths a new set of control adjustments was required: if a correct adjustment was not

TABLE 3
GROUP ERROR SCORES FOR EXPERIMENTAL CONDITIONS

Predictability	Slide Sequence	Structure Sequence	Rate of Change			
			10-second		15-second	
			1	2	1	2
Random	S.S. I.	AD	347	337	168	33
		DA	174	219	62	162
	S.S. II.	AD	278	184	229	64
		DA	242	247	257	247
Predictable	S.S. I.	AD	296	115	130	45
		DA	170	217	42	66
	S.S. II.	AD	208	203	69	199
		DA	225	260	83	164

TABLE 4
SUMMARY OF ANALYSIS OF VARIANCE:
GROUP ERROR SCORES

Source of Variance	df	Mean Square	F Ratio
1. Total	31	13,239.660	
2. Between groups	15		
A. Rate of change	1	5,029.174	20.479 ($p < .01$)
B. Predictability	1	997.507	4.062
C. Slide sequence	1	576.000	2.346
D. Structure	1	8.028	
sequence			
A × B	1	14.063	
A × C	1	693.444	2.824
A × D	1	225.000	
B × C	1	12.250	
B × D	1	.111	
C × D	1	733.507	2.987
Pooled higher order interaction	5	245.571	
3. Within groups	16		
E. Session	1	82.507	2.432
F. Structure	1	1,320.111	53.757 ($p < .05$)
E × A	1	16.764	
E × B	1	166.840	4.917
E × C	1	51.361	
F × A	1	1.361	
F × B	1	81.000	
F × C	1	253.340	7.467
E × A × B	1	291.840	8.601
E × A × C	1	14.694	
E × B × C	1	850.694	25.072 ($p < .05$)
F × A × B	1	342.250	10.087
F × A × C	1	15.340	
F × B × C	1	166.840	4.917
Pooled higher order interactions	2	33.934	

Note.—Homogeneity of the variances attributable to higher order between and within interactions, respectively, was tested by Bartlett's test (2). The hypothesis could not be rejected in either case.

made before the next slide change occurred, an error was recorded for that control. The error score for a group, then, consisted of the number of errors for the six controls for the 81 slide changes.

Table 3 presents the error scores for the experimental conditions. The results of analysis of variance of these data are presented in Table 4.

Differences for two of the three experimental variables of primary interest are significant at the .01 level. Error scores are lower for the longer (15-second) interslide interval and for the structure condition requiring the least amount of information transmission (HA). The "predictable" order of slide changes did not result in significantly lower error scores than the "random" order although the differences were in the expected direction. The interactions between task variables or be-

tween structure and task variables were not significant.

Although not presented here, a further analysis of scores for the individual control adjustments and periods indicated consistency of group performance during periods as well as over-all improvement in performance over periods. Similar results had been obtained in the previous study (7).

Task Conditions

Examination of errors for the four task conditions indicates an ordering of scores over conditions. The 10-second random condition appears to be the most "difficult," with the 10-second predictable, 15-second random, and 15-second predictable representing decreasing levels of "difficulty." It is apparent that the 10-second rate of change necessitates the processing of more information per unit time than the 15-second rate and that the predictable condition presents cues as to the sequence of information changes not presented under the random condition. Thus, both the rate of change and predictability variables modify the number of equally likely information states per unit time among which the group must discriminate.

A method of measuring the "information" or "uncertainty" of a source is available from information theory (5, 6). Applied to the present variables, the average amount of information presented a group in bits per minute for the four conditions is as follows:³ 10-second

³There were two instruments per slide and three possible readings on each instrument, thus 3^2 equally probable information alternatives for each slide. The information (in bits), per slide is, therefore, equal to the $\log_2 3^2$. At the 15-second rate of change the information in bits per minute would be four times this value, while for the 10-second rate it would be six times this value. When slide changes occur in a random order however, the equiprobable states at each slide change consist of the number of alternatives for each of the slides (3^2) multiplied by three, since each slide has an equal chance of being presented. The information in bits per slide change is thus equal to the $\log_2 3 \times 3^2$. For the predictable sequence all 27 possible alternatives are available for the first change, 2×3^2 for the second change, and 3^2 for the third and all subsequent changes since the order of slide changes is established after three changes. The information in bits per minute for the 10- and 15-second rates of change for the random order than are $6 \log_2 27$ and $4 \log_2 27$ respectively. The information in bits per minute for the 10- and 15-second rates for the predictable order are approximately $6 \log_2 9$ and $4 \log_2 9$ respectively. In the latter

TABLE 5

MEAN GROUP ERROR SCORES FOR STRUCTURE AND TASK CONDITIONS

	Rate of Change			
	10-second		15-second	
	Random	Predictable	Random	Predictable
Low autonomy	272.75	245.25	201.50	107.25
High autonomy	234.25	178.25	104.00	92.25

random, 24.12; 10-second predictable, 18.96; 15-second random, 16.08; 15-second predictable, 12.64. It is evident that error scores are an increasing monotonic function of the amount of information per unit time presented the group.

Previous results obtained with individuals have indicated a relationship of essentially similar form between information rate and performance criteria (3, 4). However, even at the minimum input rate for our groups (12 bits per minute) a considerable number of errors were made. For individuals performing simple discrimination tasks a capacity of approximately 5 to 20 times this rate has been established (5, 6). The discrepancy presumably reflects inefficiency in the systems arrived at by the groups for processing the presented information, and probably would not be so large if "optimum" procedures for routing information were used. However, further study on this point is needed.

Structure Conditions and "Linkage Types"

Table 5 presents the error scores for the LA and HA structures for the four task conditions. Error scores for both structures increase monotonically from the 15-second predictable treatment to the 10-second random treatment and thus, according to the above analysis, increase as a function of the amount of information presented per unit time. For both structures errors are essentially a negatively accelerating function of amount of information. However, there is an indication that differences between the two structures are maximal for the two

case, the information values were actually computed by averaging the information per slide change over the 81 slide changes since, as indicated above, the number of equiprobable alternatives is not equal to 3^2 for the first two changes.

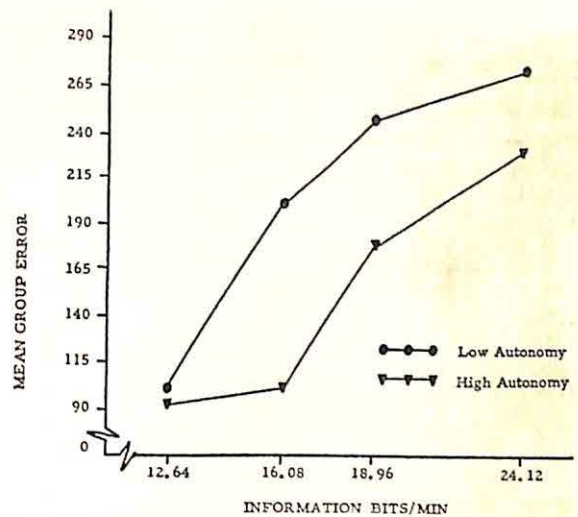


FIG. 1. RELATIONSHIPS BETWEEN MEAN GROUP ERROR SCORES AND INFORMATION RATE FOR THE TWO EXPERIMENTAL STRUCTURES

intermediate values of input uncertainty (Fig. 1).

The lesser requirements for coordination in the HA structure would lead one to expect that the relative advantage of HA over LA would increase as the amount of information presented to a group increased. Examination of the results for specific controls may shed some light on the unexpected relationships. The previous study (7) had indicated that much of the variance attributable to structure could be accounted for in terms of "linkage types."

Classification of controls in terms of whether information was directly available or relayed and whether it was located at a single source or dispersed resulted in four qualitatively difference linkage types. The four types are: controls for which the two items of informa-

TABLE 6
MEAN ERROR SCORES FOR FOUR TASK CONDITIONS BY LINKAGE TYPE

	Linkage	Rate of Change			
		10-second		15-second	
		Random	Predictable	Random	Predictable
High autonomy	Sole source—direct	119.33	98.66	43.66	35.00
	Dispersed source—partially relayed	193.00	139.00	95.00	88.00
Low autonomy	Sole source—relayed	145.00	142.00	113.33	33.33
	Dispersed source—relayed	218.66	189.66	155.33	109.66

TABLE 7

SUMMARY OF ANALYSIS OF VARIANCE: LINKAGE-TYPE
ERROR SCORES

Source of Variance	df	Mean Square	F Ratio
1. Linkage type	3	18,590	22.518 ($p < .001$)
A. Sole vs. dispersed	1	39,331	47.642 ($p < .001$)
B. Structure	1	16,354	19.809 ($p < .01$)
A \times B	1	86	
2. Source uncertainty	3	24,268	29.396 ($p < .001$)
3. Linkage type \times uncertainty	9	832	
4. Within (error)	32	826	

tion required were directly available (sole source—direct); controls for which the two items of information required were available from one of the other group members (sole source—relayed); controls for which one item of information was directly available and one item was available from another group member (dispersed source—partially relayed); controls for which one item was available from each of the other group members (dispersed source—totally relayed).

Table 6 presents the error scores for the four task treatments by linkage type. The results of analysis of variance, shown in Table 7, indicate that variance attributable to linkage types is significant as is that attributable to whether the information is dispersed or available at a single source. Structure and "uncertainty of the input" are also highly

significant. The interaction between linkage types and "uncertainty" is not significant.

The significance of "sole" vs. "dispersed" source indicates why the relative superiority of HA over LA does not increase as the uncertainty of the input to the group increases. Errors for sole-source controls appear to rise rapidly at first and then level off, whereas errors for dispersed-source controls continue to increase linearly as uncertainty increases (see Fig. 2). Since both LA and HA include sole-source controls as well as dispersed-source controls, mean errors for these structures tend to exhibit similar relationships with "uncertainty."

In terms of decreasing total error scores, the linkage types order themselves as follows: sole source—direct; sole source—relayed; dispersed source—partially relayed; and dispersed source—totally relayed. Some rather simple assumptions may explain this ordering. One may postulate that correct control actions are dependent on one or more of three primary processes—detection of a change in an instrument reading, communication of that change to the individual requiring the information, and translation of the information into control actions by reference to a table of operating instructions. The latter activity is common to all the controls but the two other processes are required to a differential degree according to linkage type.

Assuming that there is a certain probability of correct "detection" and "communication" and that such probabilities are independent across operators, then the probability of a correct setting for dispersed source linkages is a function of the joint probability of detection by two operators, while the probability of a correct setting for sole source linkages is a function of the probability of detection by a single operator. Similarly, linkages requiring the communication of information from two operators are dependent on the joint probability of correct communication from two operators while those requiring only a single communication are a function of the probability of correct communication from one operator. Assuming further that for all operators the probability of "detection" and "communication" are approximately the same, the joint probabilities of detection or communication reduce to the square of the probability for a single operator. Designating the probability

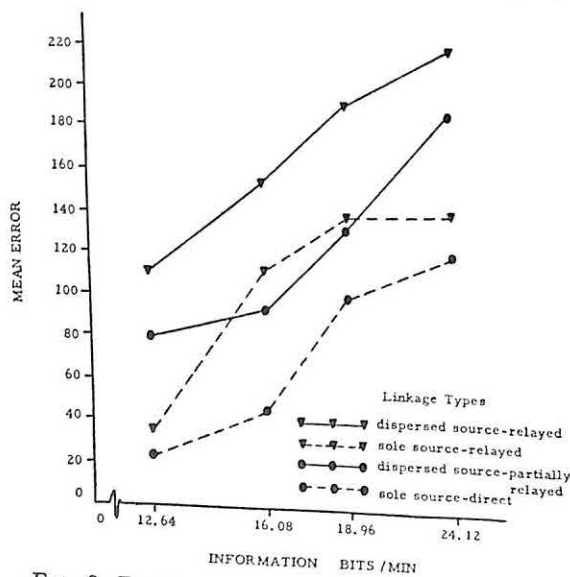


FIG. 2. RELATIONSHIPS BETWEEN MEAN ERROR SCORES AND INPUT RATE FOR THE FOUR LINKAGE TYPES

of detection by P_d and of communication by P_c , the probability of correct control actions for the four linkage types would be as follows:

1. Sole source—direct:

$$P_1 = f(P_d).$$

2. Sole source—relayed:

$$P_2 = f(P_d \cdot P_c).$$

3. Dispersed source—partially relayed:

$$P_3 = f(P_d^2 \cdot P_c).$$

4. Dispersed source—totally relayed:

$$P_4 = f(P_d^2 \cdot P_c^2).$$

It is apparent that for all values of P_d and P_c except 1, P_d^2 and P_c^2 will be less than P_d and P_c respectively, and thus that the linkage types would order themselves as was empirically found. The probability analysis, of course, implies much more than a simple ordering, but a quantitative test of the deductions must await the collection of data which would allow an estimation of the parameters. In the present data there is no way of accurately separating errors of detection from errors of communication or translation.

SUMMARY AND CONCLUSIONS

Three-man groups performed a simple "team" task involving the processing of presented instrument readings, the relaying of necessary information to individuals requiring it, and the execution of control adjustments based on relayed or directly available instrument readings. The loci of controls were varied to produce two work structures differing in the extent to which they required the transmission of information. In addition, the rate of change of the instrument readings, the predictability of the changes among individuals, and the sequence of instrument changes for an individual, were manipulated. The three task variables were arranged in a $2 \times 2 \times 2$ factorial design with two groups assigned to each cell. Each of the two groups performed under both structure conditions, the order of presentation of the conditions being counterbalanced for each cell. The performance measure used was the total number of times each control was incorrectly adjusted.

In agreement with results previously ob-

tained (7), the most difficult structure was that in which a larger proportion of information had to be relayed and, more critically, in which a larger proportion of information had to be relayed from several different sources. The importance of the latter factor was reinforced by the analysis of linkage types (in terms of whether the individual's sources of information were direct or relayed, and concentrated in a single source or dispersed) which indicated that the results for structure may be compounded of a specific linkage-type effect and a general context or difficulty effect, the latter being primarily a function of the volume of information to be relayed, the former primarily reflecting the extent of dispersion of relevant information.

Errors for both structure conditions significantly increased when the rate of change of instrument readings increased. There were no significant effects attributable to an increase in the sequential dependency between instrument changes, presumably because such dependencies were not fully perceived. An information measure applied to the task variables indicated an increasing monotonic relationship between errors and information rate. There were a considerable number of errors even at the lowest rate of approximately 12 bits per minute, a rate considerably lower than the estimates of channel capacity obtained with individuals in discrimination and reaction-time studies (e.g., 5).

This finding and the results for structure would appear to indicate that the limiting factor in the performance of the groups was not their gross information capacity. Rather the difficulty seems to lie in the inability of groups to set up an efficient system for detecting and communicating information changes. Communication problems may result from ignorance on the part of response agents as to when information bearing on their controls enters the group at some other station, and on the part of information-source persons as to the relevance of new information they receive. Detection difficulties may be a function of a response conflict generated by placing the individual in the dual role of response agent and information source. At any rate, the assumption of a high degree of dependence of control actions on both "detection" and "communication," with some further assump-

tions concerning the theoretical reliability of these processes for the various linkage types, accounts for the ordering of linkage types obtained. The interpretation is only weakly supported by the results, however, since the probability analysis implies much more than a simple ordering. Communication records are currently being obtained from which more decisive evidence may be gleaned.

REFERENCES

1. ANDREAS, B. G., & WEISS, B. *Review of research on perceptual-motor performance under varied display control relationships*. Rochester: Univer. of Rochester, May 1954. (Contract No. AF 30(602)-200, Sci. Rep. No. 2.)
2. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
3. GROSSMAN, E. R. Entropy and choice time: the effect of frequency unbalanced on choice response. *Brit. Quart. exp. Psychol.*, 1953, 5, 41-52.
4. HYMAN, R. Stimulus information as a determinant of reaction times. *J. exp. Psychol.*, 1953, 45, 188-196.
5. LUCE, R. D. *A survey of the theory of selective information and some of its behavioral applications*. New York: Columbia Univer., Bureau of Applied Social Research, June 1954. (Tech. Rep. No. 8.)
6. McMILLAN, B., GRANT, D. A., et al. *Current trends in information theory*. Pittsburgh: Univer. of Pittsburgh Press, 1954.
7. ROBY, T. B., & LANZETTA, J. T. An investigation of task performance as a function of certain aspects of work-group structure. *USAF Personnel Train. Res. Cent. Res. Rep.* submitted for publication. (Unpublished manuscript on file, Crew Research Laboratory, Randolph Air Force Base, Texas.)
8. ROBY, T. B., & LANZETTA, J. T. Work group structure, communications, and group performance. *Sociometry*, in press.
9. SHAW, M. E. Some effects of unequal distribution of information upon group performance in various communication nets. *J. abnorm. soc. Psychol.*, 1954, 49, 547-553.

Received December 27, 1955.

EFFECTS OF GLUTAMIC ACID AND SOCIAL STIMULATION IN MENTAL DEFICIENCY

GUINEVERE S. CHAMBERS AND RALPH N. ZABARENKO¹

Western Psychiatric Institute, University of Pittsburgh²

INTEREST in the effects of glutamic acid on mental functioning was aroused by a publication in 1943 (10) reporting observations made in conjunction with the use of the drug as a possible treatment in convulsive disorders. Subsequent studies on its effects on learning ability in animals (2, 23) led to extensive investigations with humans (1, 14, 18, 19, 20, 21, 22). A review of the literature reveals that interest in this area of research reached a peak in 1951-52. Since then, there has been an abrupt falling off of such publications in this country with a concomitant growth of interest in the problem in other countries.

Gadson, reviewing the literature to 1951 gives a concise presentation of the animal and human studies that yielded positive results. At the same time there have been a number of investigations, both animal (7, 9, 17) and human (4, 8, 11), where negative or doubtful results have been obtained.

With research on the wane in this area, the picture remains confused. It is difficult to compare results from one animal study to another because of basic differences in research design, types of learning being tested, number of strains of animals and finally, the difference in form and quantity of glutamic acid used. Studies with humans are even more confusing when compared. In the present study an attempt was made to overcome some of the weaknesses in other experiments by:

1. Using a large sample of institutionalized defectives representing a narrow range of intellectual level.

2. Maintaining 24-hour-a-day supervision of their environment and medication.

3. Controlling accuracy of medication by the use of tablets administered by trained personnel.³

4. Using comparable tests throughout the experiment and avoiding the introduction of nonexperimental variables.⁴

5. Including a placebo to run concurrently with the glutamic acid therapy.

6. Avoiding a "halo" effect by not revealing which tablet is the drug and which the placebo.

7. Describing in detail the environmental surroundings which conceivably could have had a bearing on results.

METHOD

Subjects. The 50 male and 8 female Ss used in the experiment were residents of Polk State School, an institution for the care and training of mental defectives. On selection, IQs ranged from 20 to 66 with a mean of 54. To permit evaluation in terms of differential effects which might be attributable to differences in mental growth rates, two distinct age groups were represented. An immature group consisting of 32 Ss ranged in age from 5 to 11. The remaining 26 Ss constituted a mature group with an age range of 19 to 35. Previous testing at the institution, in many cases covering a ten-year period, indicated that the IQs at this intelligence level were quite stable.

In the hope of isolating differential reaction to the drug, Ss were selected from three etiological categories: (a) familial mental defect; (b) defect due to abnormal developmental conditions; (c) postnatal brain damage. Selection of Ss was aimed at avoiding cases with psychoses or convulsive disorders. However, during the course of the experiment one mature S gave evidence of bizarre ideation and 2 immature Ss displayed convulsive behavior, but all were retained throughout the study. No mongoloids or phenylpyruvic oligophrenics were included.

Experimental environments. The Ss were treated in two distinctly different environments in an effort to answer the question asked of other studies in which increase in intellectual functioning is under scrutiny: How much noted change is the result of the experimental variable and how much to an increase in social opportunities and a change in self concept?

³ Zimmerman *et al.* (23) comment on the difficulty of getting parents to administer the drug faithfully.

⁴ *Ibid.* These workers include cases who had been receiving psychotherapy and others who were being intensively tutored.

¹ Formerly Associate Research Psychiatrist, Western Psychiatric Institute, Pittsburgh, and presently, Instructor, University of Pittsburgh School of Medicine.

² The authors wish to express their appreciation to: Dr. G. B. Pearson, Former Director of the Institute; Dr. Gale H. Walker, Superintendent of Polk State School; Dr. R. A. Patton, Research Coordinator for the project; Dr. A. W. Bendig, who gave statistical advice; Mrs. Frances Williams, Former Directress of Nurses; Mrs. Helen Galloway and Miss Christine Haines, Dieticians; and especially to nurses, attendants, and other personnel too numerous to be given personal acknowledgment.

Twenty-seven male Ss (16 immature, 11 mature) were transferred to the Western Psychiatric Institute and Clinic in Pittsburgh (WPI). The other 31 Ss remained in their accustomed environment at Polk. All Ss were selected by the medical director of the project from the imbecile and low-grade moron groups for comparability of institutional history and similarity of defect within each subgroup, and were randomly assigned to the WPI and Polk groups. It developed that there was a significant difference in mean IQ between these two groups (44.09 and 34.59 for the Polk and the WPI groups respectively).⁵

Polk School, a well-run institution, is located in a rural area that allows considerable freedom of movement for patients. Turnover in personnel is low. In many instances attendants are third generation employees who have taken on family roles in relation to the patients. Assignment of living quarters in homelike cottages is made according to age, sex, and degree of defect. Throughout the study, the living conditions of the Polk group remained essentially the same except for those of eight small boys. To simplify the medication procedures, these Ss were transferred to a unit which they shared with approximately 90 adult males. This resulted in their receiving almost constant attention and becoming cottage favorites. All medication was administered by two graduate psychiatric nurses from WPI who resided at Polk for the duration of the experiment. In addition to the medicating, they made detailed behavioral notes but were not responsible for the routine care of the Ss.

In sharp contrast to Polk, WPI, a modern structure with hospital atmosphere, is located in the business and cultural center of a city. One floor was set aside as living quarters for the Ss. The ratio of personnel to patients was much higher in the WPI situation than at Polk. Freedom to initiate and carry out individual activities was greatly restricted at WPI because of its urban setting. The city environment, on the other hand, provided opportunity for many new experiences. Relatives made more frequent visits to patients because of the better transportation facilities in the urban area, and also, apparently, because they had renewed interests and hopes for these children about whom they had previously adopted an attitude of resignation.

Experimental schedule. Because of the added variable of change in environment for the WPI group, psychological tests were administered on five occasions in contrast to three for the Polk group. These additional testings occurred in the initial phase of the experiment to ascertain the effects of change of environment before the introduction of medication. The WPI group was tested (a) on selection; (b) immediately after transfer to the Institute; (c) at the end of a six-month adaptation period; (d) at the close of the first medication period; (e) at the end of the second medication period. Testing for the Polk group corresponds to administrations a, d, and e for the WPI group. Both at Polk and WPI the mature and immature groups were divided into two treatment subgroups. Each S served as his own control by being in the drug group for one treatment period

and in the placebo group for the other. Glutamic acid and placebo treatment ran concurrently throughout.

Medication. L+ glutamic acid made up into 0.5-gram tablets was used as the experimental drug, while a placebo consisting of starch, lactose, citric acid, and oil of peppermint was the control treatment. The drug and placebo were identical in appearance and very similar in taste. Except for the medical director, none of the personnel knew which medication contained the glutamic acid, although it was known that a placebo was being used. Psychological test results were in the possession of the advisory committee before the psychologist knew when each group had been in the experimental or control phase.

In the early stages of the experiment, the dosage levels were increased very gradually, but when no disturbing effects were noted at maximal dosage, later groups were brought to full 40-gram intake over a period of 10 days. Each S was given daily doses of 40 grams of glutamic acid or an equivalent number of tablets of the placebo during the control period. Medication was administered at the same hours each day, three times a day. The bulk of the daily dose was given early in the day to avoid possible evening restlessness, e.g., at maximal dosage each group was receiving 39 tablets at 10 A.M., 27 at 3 P.M., and 14 at 8 P.M. The medication periods were at least 100 days in length, with somewhat longer periods for several groups necessitated by the amount of time required for travel and testing. Each S was individually medicated. In effecting the change-overs in type of medication, the total number of tablets remained the same, a few tablets of the new medication being gradually substituted for the old until full dosage was reached at the end of a 10-day period.

With rare exception, the Ss were entirely cooperative in taking the medication. The mature Ss were especially eager for their treatment, feeling that as a result they were going to be "stronger" and "smarter." The fact that these Ss recognized themselves as a small segment of the Polk population in itself lent status to their position. In the approximately 10,000 patient days of medication there were at most a half dozen instances of gastric disturbances. A relationship to medication could not be demonstrated.

Testing procedures. All psychological tests were administered by one psychologist who became known to the Ss before initiating testing to insure adequate communication and maximal cooperation.

The test battery for the mature Ss was: (a) Stanford-Binet, Forms L and M alternated; (b) Grace Arthur Point Scale of Performance Tests; (c) Rorschach; (d) Vineland Social Maturity Scale; (e) Bender-Gestalt; (f) Human Figure Drawings; (g) Test of Animistic Thinking (12); (h) Test of Concept Formation; (i) Minnesota Rate of Manipulation. The battery for the immature groups was (a) Stanford-Binet or Merrill-Palmer;⁶ (b) Vineland Social Maturity Scale.

RESULTS

Psychiatric observations. A report of the psychiatric aspects of the study is presented in

⁶ Mental ages for 4 immature Ss are from the Merrill-Palmer Scale because of inadequate language development.

⁵ More clear-cut conclusions could have been reached in regard to the two environments had this difference been anticipated.

the initial report on this research (16). Briefly, the most pertinent findings were that the Ss who were removed from their stable, familiar surroundings showed marked signs of distress both in physiological functioning and regressive social behavior, which quickly disappeared. Not only did Ss in this group regain former competence, but they also gave indications of response to the added attention by appearing more alert and developing more complicated patterns of social behavior. There was a low ceiling to the amount of social growth that could be stimulated, and they remained on this plateau for many months. Toward the end of the study they found the life at WPI restrictive and lost some of their earlier hope of becoming "smarter."

The positive changes observed in the WPI group were also seen in the Polk group, but these continued to make their appearance over a longer period than at WPI. These changes were especially notable in the group of eight small boys who had been transferred to the adult cottage.

Psychological test results. Considering first the effect of glutamic acid as measured by intelligence test scores, analysis of variance revealed no significant differences in the means for any group between the drug and placebo treatments. The data are given in Table 1. An analysis was made of the Stanford-Binet data for Ss who gained 10 or more months in mental age during the treatment period. The cumulated gains were broken down into those achieved when the drug was the first medication and those occurring when the placebo was administered first. The largest gain made by any one individual during the treatment

TABLE 1
COMPARISON OF PERFORMANCE IN MENTAL AGE
MONTHS OF THE POLK AND WPI GROUPS
UNDER GLUTAMIC ACID AND
PLACEBO TREATMENT

Source of Variation	df	Mean Square	F Ratio
Between Groups	3	6591.887	19.570**
Ss in Same Group	54	336.844	
Between Treatments	1	42.241	2.134
Interaction: Treatment × Groups	3	2.346	<1.0
Interaction: Pooled Ss × Treatment	54	19.741	
Total	115		

** $p < .001$.

TABLE 2
COMPARISON OF PERFORMANCE IN MENTAL AGE
MONTHS OF THE WPI GROUP TESTED
UNDER THREE ENVIRONMENTAL
CONDITIONS

Source of Variation	df	Mean Square	F Ratio
Between Maturity Levels	1	9508.94	23.73**
Between Ss in Same Group	25	400.69	
Between Conditions	2	128.59	12.57*
Interaction: Conditions × Maturity Level	2	8.95	<1.0
Interaction: Pooled Ss × Conditions	50	10.23	
Total	80		

* $p < .01$.

** $p < .001$.

period was 19 months. The difference between the drug and placebo gains again was not significant nor related to the order of medication.

The data from the first three testings on the Stanford-Binet for the WPI group were examined to assess the effects of transfer and an adaption period. The differences in these means were significant (.01 level, Table 2). Applying the t test, the source of the difference could be attributed to gains accrued during the socialization period, between transfer ($M = 48.5$) and the end of the six months of adaptation ($M = 51.33$), the mature and immature Ss having contributed equally to the total gain which was significant at the .001 level. The difference between initial testing performance ($M = 47.0$) and that following transfer was not significant.

Participation in the experiment was accompanied by significant differences in Stanford-Binet scores in all groups (.001 level) even though there was no difference found between the medication and placebo effects. In consideration of the 16-month time lapse and the inclusion of immature Ss who constitute over half the total number, and who would be expected to be developing, even at a retarded rate, the over-all change was not impressive.

Considering total groups, immature Ss made greater gains than mature ones (.05 level), and the Polk group made greater mean gains than the WPI group (.01 level). Since there was a significant difference between the WPI and Polk groups at the time of selection, a further analysis was made in an effort to determine whether ability to respond to environ-

TABLE 3
MONTHS MENTAL AGE CHANGE OF EIGHT
IMMATURE MALE SUBJECTS

Case Number	Glutamic Acid	Placebo	Over-all Gain
43	5	1	6
46	0	6	6
47	1	3	4
50	2	13	15
51	13	1	14
53	8	3	11
54	17	1	18
57	5	10	15

mental stimulation may be dependent upon a minimal basic capacity. All Ss were dichotomized into those above and below IQ 45. Again, the relationship held for maturity level, with immature Ss making significantly greater gains (.05 level) than mature Ss with the younger Ss having IQs above 45 making the greatest gains.

The changes in mental age scores of the eight small boys at Polk (Table 3) are of especial interest since, with one exception, they maintained an approximately normal rate of gain in mental age over the experimental period.

While there were no differential reactions to the drug among the three diagnostic categories, there were significant differences between groups (.05 level) in terms of participation in the study. The postnatal defect group (holding location constant) made significantly greater gains than did the familial or developmental defect groups.

Analysis of the mean Grace-Arthur scores revealed no significant differences in performance for any group or aspect of the study. This is rather surprising in that the tasks involved would seem to be more susceptible to practice effect than would the verbal material.

There were changes in the Rorschach performances which are reported elsewhere (3). These changes could not be attributed to the administration of glutamic acid. Changes in the Human Figure Drawings closely paralleled those noted in the Rorschach and appeared to be related to greater familiarity with the task or possibly some enhancement in self concept.

No changes were demonstrable on any other tests. Even though Ss did develop new skills such as bed-making and self-care at the table, these were at a sufficiently uncomplicated

level that they made little change in Vineland Quotients. Most of the gain occurred in areas where formerly there had been no opportunity to develop these skills.

DISCUSSION AND CONCLUSIONS

Since no tests applied to the data yielded significance between drug and placebo medication, it must be concluded that for these Ss glutamic acid did not improve mental functioning. For many Ss the test scores following the glutamic acid treatment were numerically the lowest. It was the examiner's opinion, based on behavioral notes made at each testing session, that under the drug therapy there was some lessening in ability to inhibit responses which led to impulsive errors, especially on the motor tasks.

It is of interest that the mental-age scores of these patients did not change immediately after transfer when their environment was drastically altered. The most plausible explanation for this appears to be that after an initial reaction to change (nausea, sleeplessness, crying, and loss of previous skills in self-care), they quickly returned to long established patterns of response. The finding does not appear to be attributable to insensitivity in the clinical instruments used, since mental-age scores that had been remarkably stable over many years did change significantly after a treatment-free adaptation period. Studies on the effects of glutamic acid which did not include this control period could have reported significant gains which were the product of the uncontrolled variable—increased attention—rather than the effect of the drug. It cannot be overemphasized that any environmental manipulation of a mental defective is an important variable in an experiment.

The research was designed to keep the Polk environment as similar to pre-experimental conditions as possible. Yet, the awareness on the part of the attendants that these Ss were participating in the experiment actually produced considerable change in this environment. As these people were being periodically queried about changes noted in the Ss, they became more aware of the patients as individuals and were quite hopeful that their particular charges would be improved with the drug. While conclusions from this research on institutionalized Ss are not directly applicable to studies of non-

institutionalized defectives, it is likely that in those situations, too, the Ss were receiving added attention and were responding to new hopes held for them by parents. Sarason's (13) survey of the literature on the effects of environmental stimulation adds support to this hypothesis.

Because more of the Polk Ss initially were at the higher end of this narrow intellectual range, the differential effects of the two environments are somewhat clouded. The important fact is that *both* groups received added attention and both groups made over-all gains, not related to medication. The finding of the greater gain having occurred at Polk was anticipated by directing personnel. Individuals accustomed to handling mental defectives have long been aware of the fostering effects of uninterrupted routine on their comfort and level of performance. Although the WPI Ss obviously enjoyed some novel experience such as a trip to the zoo, the changes in scheduling necessitated to provide these activities always carried with them considerable restlessness and distress. Frequent changes in personnel at WPI also seemed to make this a less secure environment than was offered at Polk.

It is not surprising that the younger children were more responsive to the stimulation provided by the experiment than were the adults. It seems logical that if a spurt of development can be induced, it is more likely to occur in an organism in the process of developing than in one where an apparent ceiling has been previously attained. When the scores of the mature Ss from childhood through their institutional life were plotted, a peak was reached typically at age 10 to 12 followed by a gradual decline to a long-standing plateau. The changes that occurred during the experiment did not exceed the highest scores on early tests, so that rather than representing increased capacity, they seemed merely to reflect higher motivation to utilize native potential. If such was the case, differentials in gains may possibly be a function of the strength of individual desire to improve.

While it is impossible from these data to account for the marked differences in over-all reaction to the experiment, the most promising lead comes from consideration of the group of boys at Polk who had, prior to the experiment, been developing at the rate of 25 to 50 per cent

of normal expectation. It was within this subgroup that the greatest and most consistent gains occurred. These children had an especially enriched environment for the duration of the study. Possibly awareness of the concern of others coupled with "tender loving care" may account for over-all gains reported in this study as well as in other studies where positive findings on small samples have been attributed to glutamic acid.

SUMMARY

1. L+ glutamic acid in this study did not produce significant improvement in mental functioning.

2. Abruptly changing the environment of institutionalized mental defectives did not affect mean mental-age scores as measured by the Stanford-Binet Intelligence Scale and Grace - Arthur Point Scale of Performance Tests.

3. Increased personal attention enhanced mental functioning.

4. A significantly greater gain was made in functioning of mental defectives who were given added attention but with minimal disruption of the familiar environment.

5. Immature Ss with IQs of 45 or above in this study made the most dramatic changes in mental functioning.

6. Subjects diagnosed as being defective as the result of postnatal insult made significantly greater gains than did Ss diagnosed as being familial defectives or having had abnormal prenatal development.

REFERENCES

1. ALBERT, K., HOCH, P., & WAELSCH, H. Preliminary report on the effect of glutamic acid administration in mentally retarded subjects. *J. nerv. ment. Dis.*, 1946, 104, 263-274.
2. ALBERT, K. E., & WARDEN, C. J. Level of performance in the white rat. *Science*, 1944, 100, 476.
3. CHAMBERS, G., & HAMLIN, R. Rorschach "inner life" capacity of imbeciles under varied conditions. *Amer. J. ment. Def.*, in press.
4. ELLSON, D. G., FULLER, P. R., & URMSTON, R. The influence of glutamic acid on test performance. *Science*, 1951, 112, 248-250.
5. GADSON, E. J. Glutamic acid and mental deficiency—A review. *Amer. J. ment. Def.*, 1951, 55, 521-528.
6. HARNEY, SISTER MAUREEN. *Some psychological and physical characteristics of retarded girls before and following treatment with glutamic acid*. Washington, D. C.: Catholic Univer. of America Press, 1950.

7. MARX, M. H. Effects of supranormal glutamic acid on maze learning. *J. comp. physiol. Psychol.*, 1948, **41**, 82-92.
8. McCULLOCH, T. L. The effect of glutamic acid feeding on cognitive abilities of institutionalized mental defectives. *Amer. J. ment. Def.*, 1950, **55**, 117-122.
9. PILGRIM, F. J., ZABARENKO, L. M., & PATTON, R. A. The role of amino acid supplementation and dietary protein level in serial learning performance of rats. *J. comp. Physiol.*, 1951, **44**, 26-36.
10. PRICE, J. C., WAELSCH, H., & PUTNAM, T. J. Glutamic acid hydrochloride in the treatment of petit mal and psychomotor seizures. *J. Amer. Med. Ass.*, 1943, **122**, 1153-1156.
11. QUINN, K. V., & DURLING, D. New experiment in glutamic acid therapy; 24 cases classified as mental deficiency, undifferentiated, treated with glutamic acid for 6 months. *Amer. J. ment. Def.*, 1950, **55**, 227-234.
12. RUSSELL, R. W., & DENNIS, W. Studies in Animism: I A standardized procedure for the investigation of animism. *J. genet. Psychol.*, 1939, **55**, 389-400.
13. SARASON, S. B. Psychological problems in mental deficiency. New York: Harper, 1949.
14. WAELSCH, H. A biochemical consideration of mental deficiency. The role of glutamic acid. *Amer. J. ment. Def.*, 1948, **52**, 305-313.
15. ZABARENKO, L. M., PILGRIM, F. J., & PATTON, R. A. The effect of glutamic acid supplemental on problem solving of the instrumental conditioning types. *J. comp. physiol. Psychol.*, 1951, **44**, 126-133.
16. ZABARENKO, R. N., & CHAMBERS, G. S. An evaluation of glutamic acid in mental deficiency. *Amer. J. Psychiat.*, 1952, **108**, 881-887.
17. ZIMMERMAN, F. T. Glutamic acid treatment of mental retardation. *Quart. Rev. Psychiat. Neurol.*, 1949, **4**, 263-269.
18. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. A group study of the effect of glutamic acid upon the mental functioning in children and adolescents. *Psychosom. Med.*, 1947, **9**, 175-183.
19. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. Effect of glutamic acid on mental functioning in children and in adolescents. *Arch. Neurol. Psychiat.*, 1946, **56**, 489-506.
20. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. Effect of glutamic acid on the intelligence of patients with Mongolism. *Arch. Neurol. Psychiat.*, 1949, **61**, 275-287.
21. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. The ceiling effect of glutamic acid upon intelligence in children and in adolescents. *Amer. J. Psychiat.*, 1948, **104**, 593-599.
22. ZIMMERMAN, F. T., BURGEMEISTER, B. B., & PUTNAM, T. J. The effect of glutamic acid upon the mental and physical growth of Mongols. *Amer. J. Psychiat.*, 1949, **105**, 661-668.
23. ZIMMERMAN, F., & ROSS, S. Effect of glutamic acid and other amino acids on maze learning in the white rat. *Arch. Neurol. Psychiat.*, 1944, **51**, 446-451.

Received September 21, 1955.

RETENTION OF AFFECTIVELY TONED VERBAL MATERIAL BY NORMALS AND NEUROTICS¹

SAMUEL F. KLUGMAN

VA Hospital, Coatesville, Pennsylvania

FOLLOWING Freud's attempt to relate forgetting to repression, a relatively large number of experiments have been carried out to study this relationship. An examination of these studies reveals many ambiguous results and weaknesses in procedure which have been ably summarized (1, 5, 11, 13, 19).

In this study, an attempt has been made to overcome the following methodological weaknesses found in earlier investigations: failure to equate for familiarity of the verbal material to be retained, insufficient number of Ss, inadequate control of overlearning and of early mastery of the learned material, neglect of intensity factors, and use of isolated words to convey emotional ideas. Furthermore, all of the quantitative studies, except Sharp's (17) which supported and Kott's (9) which failed to support Freud's theory of repression, dealt only with normal Ss. If emotional factors have a significant effect upon memory, then certainly the emotional stability of the S should be considered. In addition, "Each S must be treated with individual attention to his particular complexes . . ." (16, p. 119).

Freud's theory of repression and its relation to conversion symptoms may be related to levels of ego strength as found in normals and neurotics in such a way as to permit the prediction of differential retention between these groups in terms of quality (pleasantness-unpleasantness) and intensity (strong-mild) variables. According to Freud (3), an hysterical conversion symptom is due to a displaced affect that was part of an original traumatic incident, the recall of which would be painful to the individual. The ideational part of the original incident is subject to repression, a process serving "*the function of rejecting and keeping*

something out of consciousness" (p. 86, italics his). At a later point he states that, for neurotic patients, repression makes it "possible to bring about a total disappearance of the charge of affect" (p. 94). For this reason, in the present study, only those neurotics were used who gave evidence of conversion symptoms. It was felt that a population of this type, if Freud is correct, would yield the most definitive results since the Ss may be conceived as typically employing repression as a defense mechanism.²

Hypotheses

The general hypothesis of this study is that normal and neurotic Ss are differentially affected in their retention of learned verbal material by its affective quality and its intensity. Verbal material was employed because, according to Freud, "repression denies to the rejected idea . . . translation of the idea into words . . . The idea which is not put into words . . . then remains in the unconscious in a state of repression" (3, p. 134).

Freud further argued that, "From any operation which might arouse unpleasantness ('pain') mental activity draws back (repression)" (3, p. 14), and "The ego pursues pleasure and seeks to avoid unpleasure" (4, p. 16). In

¹ This paper is based on a dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy, in the Department of Psychology of the University of Pennsylvania. The writer wishes to express his sincere appreciation to the dissertation adviser, Dr. Julius Wishner, and to the other members of the committee, Dr. Francis W. Irwin and Dr. Miles S. Murphy, for the invaluable assistance so generously given throughout the experimental investigation.

² One effect of this procedure was to restrict the number of neurotics eligible for this experiment, making it impracticable to match the two subgroups. According to Edwards (2, p. 355), analysis of covariance is "applicable to those situations where the matching of groups is not feasible prior to the assignment of the Ss of the experimental conditions, but where some measure of initial performance may be obtained after the assignment." Furthermore, as Lindquist points out, even if the means of the initial scores for the treatment groups are not different, analysis of covariance is desirable. "Assuming some correlation between X and Y, the within-groups variance of the adjusted measures would be less than that of the unadjusted measures, so that the precision of the experiment would be increased" (10, p. 325). Following Kott (9), it was expected that a high relationship existed between learning and retention. For these reasons, analysis of covariance was used in the experiment, and both measures of retention, recall and relearning, were adjusted for initial learning scores.

terms of retention, one can derive the following hypotheses:

1. Both normals and neurotics retain more pleasant than unpleasant material.

Freud's first statement, however, also implies that the weaker ego of the neurotic resorts more readily to repression when dealing with personally meaningful unpleasant material than does the stronger ego of the normal. It follows that,

2. Normals retain more personally relevant unpleasant material than do neurotics.

In his discussion of repression, Freud has also stated, "As soon as an idea which is fundamentally offensive exceeds a certain degree of strength, the conflict (with conscious control) takes on actuality and it is precisely activation of the idea that leads to its repression" (3, p. 90). Although it may be inferred that Freud was referring only to unpleasant ideas, any strongly intense material may be generally regarded by the neurotic as "offensive." Clinical observation and Rorschach color theory (8, p. 390) support the contention that neurotics cannot properly handle affective situations and, therefore, tend to avoid highly emotional involvement. One may then hypothesize that,

3. Neurotics retain more mildly toned than strongly toned material.

Since the weaker ego of the neurotic evokes a greater need to avoid the threat of the possible appearance of a "fundamentally offensive" idea into consciousness than does the ego of the normal, it may be finally hypothesized that,

4. Normals retain more personally relevant strongly toned material than do neurotics.

METHOD

Experimental Design

Sixty male normals, World War II veterans (never treated psychiatrically), were compared with sixty male neurotics (currently being treated at a Veterans Administration Mental Hygiene Clinic) for the four possible combinations of emotional intensity (strong-mild) and quality (pleasant-unpleasant) values. Thus, Ss in each group were randomly assigned to one of four subgroups. Two subgroups, one of them neurotic and one normal, learned, recalled, and relearned a set of strongly pleasant phrases; two other subgroups did the same with mildly pleasant phrases; two other subgroups dealt with mildly unpleasant phrases; and two with strongly unpleasant phrases. The arrangement of the obtained data in Table 1 illustrates the design of the experiment.

Subjects

The neurotics were selected by their therapists, who were instructed to choose Ss who had attended high school, were between the ages of 25 and 45, and had some conversion symptom. The normals were obtained from veterans' schools, a neighborhood men's club, and a fraternal lodge in such a way as to meet the same age and educational standards as those of the neurotics. The mean age for the normals was 30.70, *SD* 4.37; for the neurotics, the mean was 29.77, *SD* 3.52. The mean educational level for the neurotics was 11.45 grades, *SD* 2.56; and 11.02, *SD* 2.50, for the normals. The neurotics were diagnosed according to instructions included in VA Technical Bulletin 10A-78, entitled *Nomenclature of Psychiatric Disorders and Reactions*. There were 18 cases of anxiety reaction, 17 of conversion reaction, and 25 of somatization reaction.

The Phrase Construction Technique

The phrase construction technique was developed to obtain personally meaningful material from each S. A chart, made up of four pages, was used. The first page contained the directions for constructing either pleasant or unpleasant phrases, the second had a scale of feeling tone with location points for mild and strong descriptions, the third a list of 365 adjectives, and the fourth a list of 372 nouns. The words were selected from the *Teacher's Word Book of 30,000 Words* (18) by three psychologists working independently, who were instructed to select words (usually taught in Grades I through IV) that might be related to social, vocational, familial, love, and health areas. Following are some examples of phrases which were constructed by the Ss according to instructions to be described below: happy mother, excellent job, cold lunch, dirty home, cruel father, faithful wife, happy children, and crying baby. This plan provided for (a) permitting S to select materials personally meaningful to him, (b) equating for familiarity and difficulty, and (c) use of more than isolated words in an attempt to obtain more meaningful ideational content. Thus, Sears' concern that "each subject must be treated with individual attention to his particular complexes" (16, p. 119) appears to have been met to some extent.

The degree of consistency in the ratings of the phrases for emotional intensity during the one-week interval was determined. Tetrachoric correlation coefficients ranged from .77 to .84 for consistency in rating for the group as a whole and for the separate subgroups. Reliability seemed sufficient to justify further study of the experimental phrases.

Learning and Retention Procedures

First session. The following description of the first normal S, who worked with mildly unpleasant material, is illustrative of the experimental method. He was told to construct 24 phrases which he considered unpleasant according to the directions contained in the chart, half of them of strong and half of mild intensity. Since he was to work with mildly unpleasant material, only his mild phrases were utilized for the experiment proper.³

³ Similar approaches were taken by Ss working with strongly unpleasant, mildly pleasant, and strongly

While *E* then printed by hand a set of learning cards, *S* read some irrelevant material⁴ which was intended to minimize any rehearsal effects. Following the printing of the learning cards and the preparation of the recording form, *S* was requested to learn the phrases, each of which was associated with a nonsense syllable. The learning procedure was carried out by a combination of the paired associate and prompting methods. The same nonsense syllables, all of which had a zero association value according to Glaze's (6) classification, were used for all *Ss*.

The phrases were learned to the criterion of two successive perfect repetitions in the following manner: first, *S* studied each association of a phrase with a nonsense syllable for ten seconds; second, the cards were shuffled and placed in an exposure box (with an opening to the left so that only the syllable of the top card was exposed); third, *S* tried to recall the phrase in five seconds. If he recalled incorrectly or not at all within that time, he was allowed to study the card for five seconds more. After each trial (consisting of going through the set of cards once), the cards were shuffled in a manner suggested by Osgood (12) to prevent serial learning. As soon as *S* had learned a card to the criterion, it was removed from the set to prevent overlearning. After every fifth trial and at the conclusion of the learning procedure, *S* read once aloud each of the cards that had so far been removed from the set. These review trials were given in order to equalize the degree of mastery over those items learned early with those learned more recently. The score was the number of trials necessary to learn all associations to the criterion.

In compliance with a request made by several school directors, a limit was arbitrarily set so as not to keep *Ss* away from their classrooms too long. If *S* took more than twenty trials to learn, he was removed from the experiment and the next *S* took his place.

Between this session and the next, *E* typed *S*'s original phrases, arranged in random order, on a mimeographed form containing directions for reclassification of the phrases at the conclusion of the next session.

Second session. During the second session, which occurred exactly one week later, at the same hour, *S* was required to recall as many phrases as he could from a set of cards that contained only the nonsense syllables. He was allowed 30 seconds in which to recall the associated phrase. His score was the number of words correctly recalled. Following this, his original set of learning cards was placed in the exposure box, and *S* was asked to relearn the associations. Except for the

fact that he did not study each card initially for ten seconds, the procedure was exactly the same for relearning as for learning. Finally, *S* was presented with the reclassification form containing his 24 original phrases, randomly arranged, and directed to indicate his present attitude as to whether each phrase was mild or strong in emotional intensity.

RESULTS

Learning

The mean learning scores reported in Table 1 represent the number of trials necessary to learn the phrases to the point of two consecutive perfect repetitions. It is apparent that the two groups, in general, were alike in their ability to learn this type of material. An analysis of variance of these scores indicated no significant differences among the variables singly or in interaction. In the case of the triple interaction, a significant inverse *F* ratio at the .05 level was obtained, but as Edwards (2, p. 228) points out, such a ratio has "no reasonable interpretation other than that (it is) the occasionally significant value which (is) to be expected from random sampling." Bartlett's test for homogeneity of variance was nonsignificant, and inspection of the distributions in the separate cells showed no unusual skewness.

TABLE 1
LEARNING AND RETENTION MEANS FOR NEUROTIC AND NORMAL GROUPS*

	Strongly Pleasant	Mildly Pleasant	Mildly Unpleasant	Strongly Unpleasant
A. Neurotics				
Learning				
Mean	13.40	12.83	13.13	13.73
SD	2.53	2.92	3.36	2.64
Recall				
Original Mean	9.33	10.60	9.73	8.73
SD	2.10	2.06	2.28	1.71
Relearning				
Original Mean	8.40	7.40	7.73	8.67
SD	1.71	1.82	1.93	2.20
B. Normals				
Learning				
Mean	12.80	13.47	13.80	13.27
SD	2.56	2.69	2.40	2.65
Recall				
Original Mean	10.73	9.33	9.00	10.40
SD	1.79	1.64	1.67	1.78
Relearning				
Original Mean	7.53	8.53	8.00	7.27
SD	1.60	2.01	1.21	1.76

* Learning and Relearning scores: trials to learn; Recall scores: words recalled.

pleasant material. The instructions to the *Ss* during the first and second sessions, and the phrase construction chart, with separate directions for pleasant and unpleasant phrase construction, have been deposited with the American Documentation Institute. Order Document No. 5033 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting \$1.75 for microfilm or \$2.50 for photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

⁴ *Comic Relief, a Collection of Modern American Humor*, by R. W. Linscott, New York: Houghton Mifflin, 1932.

These results, revealing no differences in learning between the groups, are similar to those found by Sharp (17), Kott (9), and Schmidt (15).

Correlations Between the Various Measures

Table 2 presents all product-moment correlations found among the various variables for each of the subgroups and the total sample. All coefficients are significant at the .01 level. In no instance, as determined by z values, was there a significant difference between any pairs of coefficients. The high correlations confirmed the need for partialling out the strong effects of learning on recall and relearning scores.

Adjusted Retention Scores

The recall data were adjusted for initial learning by analysis of covariance. As shown in Table 3, the interaction between affective intensity and normal-neurotic grouping yielded an F significant at the .01 level. No other significant F s were found. The relearning data were similarly adjusted for learning scores. Significant F s were obtained for the double

TABLE 2
PRODUCT-MOMENT CORRELATIONS BETWEEN EXPERIMENTAL VARIABLES

Variables	Total	Neurotics	Normals
Learning vs. Recall*	-.68	-.68	-.68
Learning vs. Relearning	.86	.87	.85
Recall vs. Relearning*	-.71	-.75	-.70

* Fewer trials to learn or relearn related to greater number of words recalled.

TABLE 3
ANALYSIS OF RETENTION SCORES ADJUSTED FOR LEARNING SCORES

Source	df	Recall		Relearning	
		Variance	F*	Variance	F*
Pleasant-Unpleasant (P-U)	1	4.20	2.09	1.75	2.12
Strong-Mild (S-M)	1	.42	—	.14	—
Neurotic-Normal (PN-N)	1	2.00	—	1.66	2.01
P-U × S-M	1	.28	—	.01	—
P-U × PN-N	1	1.61	—	4.52	5.29
S-M × PN-N	1	29.54	14.70	10.35	12.56
P-U × S-M × PN-N	1	.15	—	.21	—
Error	111	2.01	—	.824	—

* $F_{.05} = 3.93$; $F_{.01} = 6.87$

TABLE 4
ADJUSTED MEAN DIFFERENCES BETWEEN NORMALS AND NEUROTICS FOR QUALITY AND INTENSITY VARIABLES

	Recall			Relearning		
	Adjusted Means†		t	Adjusted Means†		t
A. Pertinent to Hypotheses						
PN-P vs. PN-U	9.90	9.28	1.68	7.99	8.14	.64
N-P vs. N-U	9.94	9.80	.38	8.14	7.45	2.93
N-U vs. PN-U	9.80	9.28	1.41	7.45	8.14	2.92
PN-M vs. PN-S	10.03	9.15	2.41*	7.74	8.39	2.76*
N-S vs. PN-S	10.43	9.15	3.50**	7.57	8.39	3.49*
B. Nonpertinent						
N-S vs. N-M	10.43	9.32	3.04**	7.57	8.09	2.23*
N-S vs. PN-M	10.43	10.03	1.09	7.57	7.74	.72
PN-M vs. N-M	10.03	9.32	1.95	7.74	8.09	1.51
N-M vs. PN-S	9.32	9.15	.46	8.09	8.39	1.2
N-P vs. PN-U	9.94	9.28	1.80	8.14	8.14	.0
N-P vs. PN-P	9.94	9.90	.11	8.14	7.99	.65
PN-P vs. N-U	9.90	9.80	.27	7.99	7.45	2.28*

* Significant at .05 level of confidence.

** Significant at .01 level of confidence.

† Number of words recalled.

‡ Trials to relearn.

interactions between grouping × quality and grouping × intensity.⁵

In view of these significant F s obtained from the double interactions, tests were made to determine the significance of the differences between the groups for quality and intensity variables, respectively, and for both measures of retention. These appear in Table 4.

The upper half of the table contains the data pertinent to the hypotheses of the study. The first hypothesis stated that "both normals and neurotics retain more pleasant than unpleasant material." The results for both groups fail to support this hypothesis; in fact, the relearning data for normals indicate significantly better retention of the unpleasant than of the pleasant material. The second hypothesis stated that "normals retain more personally relevant unpleasant material than

⁵ There is some question whether the present relearning data might be treated as Kott had, by adjusting the means for both learning and recall scores since these scores were themselves affected by experimental conditions. For this reason, although omitted from the body of this paper, the analysis of covariance data for relearning scores adjusted for both learning and recall scores have been deposited with the ADI (see Footnote 3) and are available for those who are interested. It may be noted that whereas the unadjusted groups × quality interaction was not significant, it became so at the .05 level when adjusted for learning scores, and significant at the .01 level when adjusted for both learning and recall scores.

do neurotics." The data, as measured by the relearning method, significantly support this hypothesis. "Neurotics retain more mildly toned than strongly toned material" was the third hypothesis and the findings support it significantly by both recall and relearning methods. Finally, the hypothesis that "normals retain more personally relevant strongly toned material than do neurotics" was similarly supported to a significant degree.

In the bottom half of the table, the data show that among normals, retention was significantly better for strongly toned than for mildly toned material. No other significant differences for any other comparisons were found except one: better retention by the relearning method of unpleasant material by the normals than pleasant material by the neurotics.

DISCUSSION

Effectiveness of Quality and Intensity Variables

More definitive data are available for intensity than for quality, in that there were more significant differences between the groups and stronger agreement between the two measures of retention. This observation is in line with Rapaport's summary statement after a review of recall experiments: "The intensity of the 'emotional factor' proved to be of more importance in its influence on memory than its quality . . ." (16, p. 94).

In addition to the general finding that intensity plays a more important role than quality, a pattern of effectiveness of intensity factors seems to emerge. Within the groups, neurotics retained mildly toned material more efficiently, whereas normals retained strongly toned material more efficiently. Between the groups, neurotics retained mildly toned material more efficiently than did the normals, and normals retained strongly toned material more efficiently than did the neurotics. In other words, at the time of remembering, the mildly toned material seemed to favor or be favored by the neurotics, while the strongly toned material seemed to favor or be favored by the normals. It may be assumed that all *SS* were affected by the emotionally toned material in that such material became more vivid as the intensity increased, resulting in better retention following the time lapse. However, when the material became too intense, a disruptive

effect appeared, which counteracted any increasing vividness and resulted in loss of retention. Thus, it seems that two factors, vividness (making for efficiency of retention) and disruptiveness (making for inefficiency) were operating here. Furthermore, along the continuum of intensity, there were different points of maximum efficiency of retention, followed by disruptive reactions by both groups.

It appears that neurotics are more sensitive to the emotional intensity of a situation than are normals; mildly toned material becomes vivid for them sooner than it does for normals. As the emotional intensity gets stronger, disruptiveness begins to interfere with the increasing vividness that usually accompanies stronger intensity. For the normals, who appear to tolerate increasing intensity more easily than neurotics, vividness increases for a longer period as a function of intensity. However, even in their case a degree of intensity is reached at which the disruptive effect becomes operative.

In regard to quality, it was found that the pleasantly toned material failed to differentiate between the two groups, while the unpleasantly toned material resulted in relatively better retention by the normal group as measured by relearning. This trend may be attributable to a greater degree of emotional disturbance aroused in the neurotics during the time lapse and to the better ability of the normals to tolerate this material.

Relevance for Freud's Theory of Repression

Experimental studies of repression are subject to three types of evaluation. First, there is the claim that experimental studies can prove or disprove the theory (19). Second, there is the contention that the theory cannot be tested experimentally at all (13, 14). Between the two extremes is the point of view represented by Sears (16). Although pessimistic, he holds out some hope that laboratory investigations may contribute toward some "refinement of the theory," the possible "addition of relevant new variables," or the introduction of "techniques that promise eventual solution."

For those who believe that Freud's repression theory can be tested by means of experiments of this nature, the results of this study apparently fail to support the theory. In his

discussion of the theory of repression, Freud restricted his observations to qualitative factors, i.e. pleasantness-unpleasantness, with greater retention of the former over the latter. In the present study, as pointed out above, it is precisely with regard to the comparisons between pleasantly toned and unpleasantly toned material that little or no significant differences were obtained. Thus, for both groups, there was no support for the thesis that pleasantly toned material is better retained than unpleasantly toned material. Gordon, using a medium other than verbal material, obtained similar results and stated, "Pleasant odors are not more likely to be recalled than the unpleasant . . . We must search for other factors than affective tone if we would understand why some impressions are recalled and others are not" (7, p. 225). According to the present results, retention appears to be based chiefly on the relationship between intensity factors and the emotional status of the Ss.

In this experiment, as in Kott's (9), differential effects in retention became manifest, not at the time of original learning, but after a lapse of one week. Sharp (17) found similar results in that differences failed to occur until after the second day. Schmidt (15) likewise found that his more anxious Ss did not differ in items correctly learned from his less anxious ones at the time of learning. One week later, differences in stimulus generalization responses occurred.

SUMMARY

The purpose of this investigation was to study the effects of affective intensity and quality of verbal material on the retention of this material by normal and neurotic male Ss, testing hypotheses arising from Freud's theory of repression. The groups did not differ significantly as to age and schooling. Each S constructed personally meaningful two-word phrases, which he associated with nonsense syllables and learned to a criterion of two consecutive perfect repetitions within twenty trials. A week later he was required to recall and relearn them. Analysis of covariance was employed, with recall and relearning scores as the final measures and the original learning scores as the initial measures.

The results and their relation to the hypotheses may be summarized thus:

1. The data failed to support the hypothesis that "both normals and neurotics retain more pleasant than unpleasant material."

2. The data support the hypothesis that "normals retain more personally relevant unpleasant material than do neurotics" when retention is measured by the relearning method.

3. The data support the hypothesis that "neurotics retain more mildly toned than strongly toned material" by both recall and relearning methods.

4. The data support the hypothesis that "normals retain more personally relevant strongly toned material than do neurotics" by both measures of retention.

5. There were no significant differences for outcomes based on single variables. Thus, there were no significant differences in the retention of pleasant or unpleasant materials or of strong or mild material, and there were no differences with respect to single variables between the normal and neurotic groups.

6. The phrase-construction technique developed for this study appears to be a suitable means for obtaining personally meaningful verbal material for use in studying memory function as influenced by affective factors.

7. If Freud's theory of repression is at all testable by measuring retention of learned pleasant-unpleasant verbal material, then the obtained data do not support his hypothesis of motivated forgetting since neither group retained more personally relevant pleasant than unpleasant material. Instead, the present data suggest that forgetting is based on differential interactions between normal-neurotic grouping and affective intensity and quality factors, especially the former.

8. In investigations dealing with the relationship between personally meaningful material and memory function, it seems highly desirable to consider not only the variables of quality and intensity but also the variable of emotional stability in the Ss.

REFERENCES

1. BARRETT, DOROTHY M. Memory in relation to hedonic tone. *Arch. Psychol.*, N. Y., 1938, No. 223. Pp. 61.
2. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
3. FREUD, S. *Collected papers*. Vol. IV. London: Hogarth, 1950.

4. FREUD, S. *An outline of psychoanalysis*. New York: Norton, 1949.
5. GILBERT, G. M. The new status of experimental studies on the relationship of feeling to memory. *Psychol. Bull.*, 1938, **35**, 26-35.
6. GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255-269.
7. GORDON, KATE. The recollection of pleasant and unpleasant odors. *J. exp. Psychol.*, 1925, **8**, 225.
8. KLOPFER, B., & KELLEY, D. M. *The Rorschach technique*. New York: World Book Co., 1942.
9. KOTT, M. G. Learning and retention of words of sexual and nonsexual meaning. *J. abnorm. soc. Psychol.*, 1955, **50**, 378-382.
10. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.
11. MELTZER, H. The present status of experimental studies on the relationship of feeling to memory. *Psychol. Rev.*, 1930, **37**, 124-139.
12. OSGOOD, C. E. Meaningful similarity and interference in learning. *J. exp. Psychol.*, 1946, **36**, 277-301.
13. RAPAPORT, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.
14. ROSENSTOCK, I. M. Perceptual aspects of repression. *J. abnorm. soc. Psychol.*, 1951, **46**, 304-315.
15. SCHMIDT, D. P. Some effects of manifest anxiety and symbolic reward and punishment on learning and stimulus generalization of complex visual material. Unpublished doctor's dissertation, Univer. of Washington, 1953.
16. SEARS, R. R. Survey of objective studies of psychoanalytic concepts. *Soc. Sci. Res. Coun. Bull.*, 1943, No. 51.
17. SHARP, AGNES A. An experimental test of Freud's doctrine of the relation of hedonic tone to memory revival. *J. exp. Psychol.*, 1938, **22**, 395-418.
18. THORNDIKE, E. L., & LORGE, I. *Teacher's word book of 30,000 words*. New York: Columbia Univer. Press, 1944.
19. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39-51.

Received October 19, 1955.

REMINISCENCE, DRIVE, AND PERSONALITY THEORY¹

H. J. EYSENCK

Institute of Psychiatry (Maudsley Hospital), University of London

SEVERAL attempts have been made in recent years to integrate personality study and learning theory. Of particular value in this connection has been the work of Mowrer (27), Dollard and Miller (4), and Spence (30, 31, 32). The reasons why these attempts cannot be regarded as wholly successful have been discussed elsewhere, and a "dynamic theory of anxiety and hysteria" (11) has been proposed, which attempts in a formal manner to relate the postulates of learning theory to an empirically founded and experimentally tested description of certain major personality variables.

This system of personality description has been reported in detail (7, 8, 9, 10). For the present purpose, it must suffice to say that two independent dimensions of personality (neuroticism and extraversion-introversion) have been identified quantitatively. These dimensions are related to classical psychiatry in the manner indicated by Jung: hysterics and psychopaths have high scores on both extraversion and neuroticism; whereas dysthymics, suffering from depression, anxiety states, or obsessional and compulsive symptoms, have high scores on both neuroticism and introversion. (It is not assumed, of course, that these two dimensions are the only ones which describe human conduct; what is maintained is simply that these are two very important dimensions, the understanding of which would make possible the description and control of a considerable proportion of human behavior.

PROBLEM

The postulate that links the dimensions of extraversion-introversion with modern learning theory and specifically that system of constructs used by Pavlov (29) and Hull (19) is called the typological postulate, and has been formulated thus:

¹ The author is indebted to the Research Fund of the Maudsley and Bethlem Royal Hospitals for generous help in making this investigation possible. He is also indebted to Mr. H. G. Jones for help on some of the theoretical points, and to Mr. Kolman Star, who carried out some of the computations.

Individuals in whom reactive inhibition is generated quickly, in whom strong reactive inhibitions are generated, and in whom reactive inhibition is dissipated slowly, are thereby predisposed to develop extraverted patterns of behavior and to develop hysterico-psychopathic disorders in case of neurotic breakdown; conversely, individuals in whom reactive inhibition has developed slowly, in whom weak reactive inhibitions are generated, and in whom reactive inhibition is dissipated quickly, are thereby predisposed to develop introverted patterns of behavior and to develop dysthymic disorders in case of neurotic breakdown" (11, p. 35).

This postulate mediates a large number of predictions in the fields of learning theory and perception. One of these predictions, relating figural aftereffect to extraversion, has been verified in a previous paper (12, 28). Another prediction, relating conditioning to introversion, has been verified by Franks for both neurotic and normal Ss (15, 16). The present paper is concerned with a third prediction, relating to the phenomenon of reminiscence. If, as is maintained with good evidence by Hull (19) and others (26), reminiscence phenomena are the product of inhibition originating in the massing of practice, then it follows from the typological postulate that *extraverts should show a higher degree of reminiscence than introverts*. The argument is probably too obvious to require much elaboration. Massed practice produces inhibition; inhibition is generated more quickly in extraverts; consequently, massed practice produces a greater amount of inhibition in the extraverted group. Reminiscence is produced by the dissipation of the accumulated inhibition during a rest period; reminiscence, therefore, is proportional to the amount of inhibition accumulated during massed practice. Extraverts who have accumulated more inhibition will therefore have more inhibition to dissipate and will consequently show greater reminiscence effects (11).

METHOD

Apparatus. The apparatus used was a traditional pursuit rotor. The turntable was 10 inches in diameter; the metal disc was $\frac{7}{10}$ inch in diameter, and the distance between the center of the turntable and the center of the disc was $3\frac{1}{4}$ inches. The Turntable revolved clockwise at a speed of 72 r.p.s.; S attempted to keep the

tip of an articulated rod in contact with the metal disc, having been given the standard instructions to move the rod in a circular manner and not to press down on the turntable. The experiment was carried out in a blacked-out room with controlled artificial illumination equal for all Ss. Recordings were made of the length of time during which the tip of the rod was in touch with the rotating disc during successive 10-second periods, an automatic device switching the recording between two recording chronometers every 10 seconds.

Procedure. There were three experimental sessions, divided from each other by 10-minute rest pauses. Each of the experimental sessions consisted of 30 successive 10-second trials, so that as far as the S was concerned he was working constantly for 5 minutes, had a break of 10 minutes, worked another 5 minutes, had another break of 10 minutes, and then worked another final 5 minutes. These conditions followed Ammons (2), who showed that this combination of massed practice and rest periods gives near optimum conditions for the emergence of reminiscence phenomena. Two breaks, rather than the more usual single break, were introduced in order to determine the reliability of the reminiscence score.

Subjects. The sample tested consisted of 50 experimentally naive male university students who were paid for their services. These students were volunteers who came to the Institute of Psychiatry for one day, during the course of which they were subjected to a number of different experiments. The pursuit rotor was applied relatively early during the day, so little generalized fatigue is likely to have been operative. No measures of intelligence were taken as the evidence tends to show that intelligence is not related to pursuit-rotor performance at IQ levels above 100. (The mean IQ level of similar groups used in previous experiments has usually been in the neighborhood of 120, with an *SD* of 10.) It is also known that correlations between intelligence on the one hand and neuroticism and extraversion on the other are very low or nonexistent. Little is known about the influence of age, but as the group was relatively homogeneous in this respect, few being below 19 or above 25, it does not seem likely that the results could have been influenced to any important extent by these variables.

Criterion tests. It is obviously crucial for the proper test of the hypothesis that a reasonably reliable and valid estimate of extraversion-introversion and neuroticism should be available. As has been argued elsewhere (9), the most desirable type of evidence comes from a battery of devices, including ratings, self-ratings, objective tests, and physiological measures. It was impossible on this occasion to apply a whole battery of this kind and a single measuring instrument was reluctantly used. This instrument was the Maudsley Personality Inventory.

This questionnaire consists of two scales containing 24 items defining extraversion-introversion and 24 items defining neuroticism, as well as buffer items and a lie scale. The extraversion-introversion scale has a corrected split-half reliability of .77; the neuroticism scale one of .88. The correlation between the two scales is $-.05$ (these figures are derived from the original standardization sample of 200 men and 200 women).

The scales were derived by means of an extensive

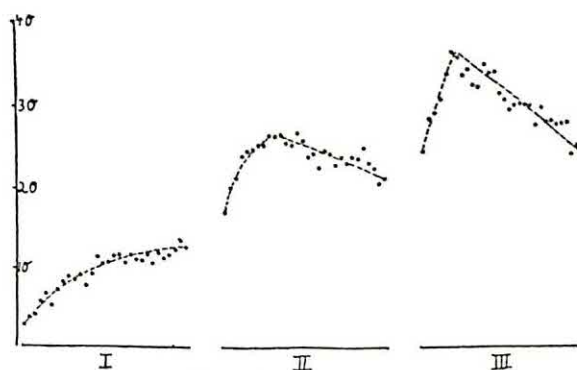


FIG. 1. MEAN PERFORMANCE ON THE PURSUIT ROTOR

The score plotted is percentage of time on target during successive 10-second trials in three 5-minute periods of massed practice, interrupted by two 10-minute rest pauses ($N = 50$).

item analysis from Guilford's STDCR and GAMIN questionnaires, and the Maudsley Medical Questionnaire. Some of the evidence relating to the development of the Maudsley Personality Inventory has been discussed elsewhere (9), but no detailed discussion of the final questionnaire has yet appeared.²

RESULTS

Reminiscence. Figure 1 shows in graphic form the average scores of the 50 Ss during the three sets of 30 trials. For ease of discussion, successive trials will be referred to by Roman numerals to indicate the first, second, and third sessions, and by Arabic numbers to indicate the trial within each experimental period. Thus, II-30 is the last (thirtieth) trial during the second period, and III-1 is the first trial during the third period. The figure shows that the main effects of experiments of this type are clearly present. These are (a) the reminiscence effect itself or the rise in performance from I-30 to II-1 and from II-30 to III-1; (b) a "warm-up" effect after the rest periods, continuing roughly from II-1 to II-10 and from III-1 to III-7; (c) a decline in performance after the peak of the "warm-up" period has been reached.

The score used here to represent individual differences in reminiscence was the difference between the average of the three last prerest trials and the first postrest trial. This measure does not take into account the practice effect of the last trial of a session, but this is probably more than compensated by not including the warm-up effect. The reason for taking the average of 3 scores as an estimate of prerest

² Copies may be obtained from the writer.

performance is simply the greater reliability of such an average as compared with a single 10-second trial score, coupled with the fact that performance appeared to have reached an asymptote. (This statement is more obviously true of I than of II.) Postrest performance cannot be measured in a similar manner because of the rapid rise of the curve (II-1 to II-10, and III-1 to III-7).³

Under these conditions, the correlation between the two reminiscence scores was .44 which, while highly significant, is clearly not very satisfactory. There are likely to be several reasons for this low reliability. The first and major one probably lies in the very high variance attaching to any single 10-second score. Chance factors, as is well known, play an enormous role in performance on the pursuit rotor. It is customary to get rid of such factors by averaging large numbers of trials or Ss. To do so is not possible in the case of individual measurement, however, and as single 10-second trials play a considerable role in the formula used here for measuring reminiscence, there is obviously a considerable amount of error variance involved in this measurement.

Another difficulty which may account for the low correlation stems from the fact that during the performance of an experiment such as this, a considerable amount of sI_R accumulates. This does not dissipate during rest, although, as argued above, some of it is extin-

guished during the so-called warm-up period. The amount of conditioned inhibition which develops is a function of (a) the amount of reactive inhibition, the dissipation of which during involuntary rest pauses is supposed to act as a reinforcement; (b) the amount of drive, which determines the level reached by reactive inhibition before these rest pauses take place, and thereby indirectly the number of rest pauses during a given period; and (c) the conditionability of the S. Depending as it does on these three variables, conditioned inhibition must be considered to vary relatively independently of reactive inhibition, thus acting as a major source of error. Kimble (20, 21, 23, 24, 25) has shown that reactive inhibition develops early and tends to remain relatively constant, whereas conditioned inhibition develops late and increases in a negatively accelerated fashion. It follows that the second measure of reminiscence (RS_2) would be far more disturbed by the error variance introduced by conditioned inhibition than would RS_1 . If this view is correct, it would explain in part the low correlation between the two, and it would also suggest that RS_1 should produce higher correlations with the indices of extraversion and neuroticism than RS_2 .

Personality correlates. The correlation between extraversion and the first reminiscence score is .29, significant at the .02 level by an appropriate one-tailed test. The correlation with the second reminiscence score is .10, which is not significant. The expectation of higher correlations with RS_1 than with RS_2 is thus borne out in the absolute magnitudes, but the difference is not significant statistically. The correlation between extraversion and the sum of the two reminiscence scores is .22, which just fails of significance at the .05 level. The correlation between neuroticism and RS_1 is .40, significant at the .01 level. That between neuroticism and RS_2 is .27. The second score is lower than the first, but not significantly so. When the two reminiscence scores are summed, they correlate .34 with neuroticism. This correlation is significant at the .02 level. (The significance of the correlations between neuroticism and the reminiscence scores was assessed by means of a two-tailed test, as no prediction was made as to direction.)

The correlation between the questionnaire measures of extraversion and neuroticism in

³ Ammons (1) has suggested a different method for the scoring of reminiscence, involving a correction for warm-up decrement. In an unpublished series of experiments, the writer has found evidence to suggest that Ammons's method of correcting for this decrement is based on a faulty theory of warm-up. The preferable alternative theory somewhat resembles that put forward by Denny *et al.* (3) to account for the fact that Ss who changed from massed to distributed practice soon reached the performance level of Ss who had been practicing under distributed conditions all the time. In brief, sI_R accumulates during the first period and, being a habit, is not dissipated during the first rest period. At the beginning of Period II, sI_R fails to be reinforced because of the absence of I_R at the beginning of the new testing period and after a long rest during which all reactive inhibition has been dissipated. Because of this failure of reinforcement, sI_R is extinguished, and consequently a steep rise in performance occurs. (The same explanation, of course, obtains in the case of the first seven trials in Period III.) If this account is accepted as a sound explanation, then clearly Ammons's method of measuring reminiscence confounds the effects of I_R and sI_R in proportions depending on the precise experimental arrangement.

this sample is quite insignificant ($r = .115$). It would therefore appear justifiable to conclude that about 25 per cent of the total variance of the reminiscence score RS_1 is predictable in terms of the two personality variables; only about 8 per cent of RS_2 is predictable in those terms. It would not be correct to say that personality factors, such as neuroticism and extraversion, *determine* the degree of reminiscence shown. Both reminiscence and personality factors are conceived of in this theory as determined by more fundamental causes, such as the inhibition/excitation balance, or the amount of autonomic drive present in a person. These are presumed to be largely innate features of the central and autonomic nervous systems respectively (13, 14) although this aspect of the theory is perhaps the one least strongly supported by direct experimental evidence. However, whether innate or acquired, these hypothetical properties of the nervous system are in this theory considered to underlie individual differences in *both* personality and learning.

Practice, decline, and warm-up. In view of the paucity of information on individual differences in various aspects of pursuit-rotor learning, it may be of interest to report some of the observed relationships. The first of these is the practice effect, defined arbitrarily as terminal score minus initial score, and obtained by subtracting the first three scores on the first 5-minute practice period from the last three scores. Theoretically, this effect would be facilitated by a strong excitatory potential and depressed by a strong inhibitory potential, so that it would be expected to correlate negatively with RS_1 and RS_2 . The correlations are, in fact, both negative ($-.48$ and $-.18$). As, however, the sum of trials 28, 29, 30 in the first period enters into both RS_1 and the score of the practice effect, the higher of these two correlations may in part be an artifact, and as the lower one is not independently significant, no conclusions can be derived from these data.

The correlations of practice score with extraversion and neuroticism are both negative ($-.103$ and $-.154$); these are both in the expected direction, but cannot be interpreted as differing significantly from zero.

A decline score was calculated for Periods II and III by subtracting the sum of the three terminal scores (II-28, II-29, and II-30; and III-29,

and III-30) from the sum of the highest III-28, three consecutive points reached in these two sessions. The hypothesis was that this decline would in part be due to reactive inhibition and should, therefore, be positively correlated with reminiscence. The two correlations in question were both positive ($.06$ and $.31$). The latter is almost significant at the $.01$ level, using a one-tailed test; the former is not significant. The reliability of this decline score is relatively low, with the correlation between the two decline scores only $.13$. It is likely that this low correlation is in part due to the differential influence of conditioned inhibition at different points on the learning curve.

A measure was taken of the warm-up effect by subtracting II-1 and III-1 respectively from the highest successive three scores obtained on Sessions II and III. These measures of warm-up show a just significant reliability, the correlation between the two being $.32$. Warm-up scores are negatively correlated with reminiscence, the warm-up in II correlating $-.69$ with RS_1 and $-.35$ with RS_2 . The warm-up in III correlated $-.30$ with RS_1 and $-.69$ with RS_2 . Again, the higher of these values are suspect as the same term enters into both sides of the correlation. Such is not the case, however, for the two smaller correlations, and it therefore seems to be established that the amount of warm-up observed is negatively related to the amount of reminiscence. The correlations between warm-up and both extraversion and neuroticism are all negative, but well below significance. It would be unsafe to assume that warm-up as measured here is related to either of these two dimensions of personality.

DISCUSSION

Little discussion appears to be necessary with respect to the relationship found between extraversion and reminiscence. The prediction made from the typological postulate was verified at an acceptable level of significance. It appears likely that a much closer relationship would be found if a more reliable and valid measure of extraversion were to be used and if a more reliable method of determining reminiscence could be devised. Experiments along these lines have indicated that shorter periods of practice (60 to 90 seconds) might give better results than the rather longer period employed in the present experiment.

Some discussion, however, appears to be necessary on the relationship between reminiscence and neuroticism. An explanation may be found along the following lines. First of all, it may be postulated with Mowrer (27), Dollard and Miller (4), Spence and Taylor (31, 32, 33), and other learning theorists that neuroticism is essentially an (autonomic) drive variable. The term usually chosen to identify this characteristic is, of course, *anxiety*. However, this term often implies not only what is psychiatrically known as anxiety state, but the whole group of neurotic disorders, so that "anxiety" and "neuroticism" are used interchangeably. This correspondence has been demonstrated by Franks (17), who found a correlation of .92 between the Maudsley Medical Questionnaire, a widely used measure of neuroticism (8), and the Taylor Anxiety Scale, a scale employed by Spence and his associates as an operational definition of anxiety. Hysterics are supposed to be psychiatrically differentiated from anxiety states in that they show relatively little anxiety. Yet hysterical *Ss* had scores on the Taylor Anxiety Scale intermediate between those of a group of anxiety patients and a group of normals (17). Consequently, it may be clarifying to talk about neuroticism as the major variable characterizing all neurotic groups, and to use the term anxiety to characterize the introverted neurotic group. There is ample evidence to justify this usage (9).

The relevance of drive (in this case autonomic drive, or neuroticism) to reminiscence follows from an extension of the Hullian theory made by Kimble (22). According to him, the amount of reactive inhibition which *S* will tolerate in a motor-learning situation is in part a function of his drive; "the greater amount of this motivation the greater amount of I_R he will tolerate" (p. 248). Kimble explains that this hypothesis follows from the fact that reactive inhibition is thought of as a negative drive or as a need to rest.

As such it acts in opposition to the needs leading to the continuation of practice. If we adopt the position that a given amount of I_R cancels a certain amount of these more positive needs, then it follows that the highly motivated *S* will continue to practice in spite of the presence of an amount of inhibition which would produce a cessation of activity in a less highly activated *S*. *The experimental prediction to which this argument leads is that the amount of reminiscence . . . will increase with an increase in motivation* (p. 248).

Both Kimble himself (22) and Wasserman (34) have provided experimental evidence in favor of this hypothesis, which may, therefore, be regarded as essentially substantiated.

The linkage of this hypothesis within learning theory to personality study is as follows. If one regards neuroticism as a drive, and if drive leads to an increase in the amount of I_R tolerated by the organism before producing a cessation of activity, then individuals with strong drives should produce greater amounts of I_R and consequently higher reminiscence scores. As this is precisely what is empirically found, one may regard this explanation as quite feasible. It would be interesting to know whether neurotics as compared with normals have higher reminiscence scores. A prediction to this effect would seem to follow from Kimble's theory and the results here reported. However, the situation may be more complicated because the drive in this connection would presumably have to be a relevant one, and it would be rather more difficult to ensure an equal relevance of autonomic drive as between hospitalized neurotics and nonhospitalized normals than it would be between *Ss* having high and low neuroticism scores respectively as in this study.

Among neurotics, one is probably on safe ground in predicting that hysterics and psychopaths will have higher reminiscence scores than anxiety states and reactive depressives (dysthymics). In terms of the MMPI and the use made of the *Hy-Pt* score devised by Eriksen (5, 6) one would probably also be correct in predicting that this score would correlate positively with reminiscence in both neurotics and normals. Thus, putting the various groups in order of expected reminiscence-score size, one has hysterics and psychopaths, followed by normal extraverts and dysthymics, with normal introverts last. It is impossible to predict with any confidence the relative positions of normal extraverts and of dysthymics; such a prediction would require a more precise knowledge of the exact degree of relationship obtaining between I_R and neuroticism and extraversion, respectively.

SUMMARY

This paper reports an experiment performed to test the hypothesis derived from learning theory, that *reminiscence on the pursuit rotor*

following massed practice should be greater for extraverts than for introverts. The relationship between reminiscence and neuroticism was also investigated. Fifty university students were employed as subjects, using pursuit-rotor learning as a measure of reminiscence and the Maudsley Personality Inventory as a measure of extraversion-introversion and neuroticism. Extraversion was found to be associated with reminiscence, according to hypothesis, and reminiscence was also found to be significantly greater for subjects high in neuroticism than for subjects low in neuroticism. These results are interpreted as indicating the fruitfulness of linking personality study with learning theory (11).

REFERENCES

1. AMMONS, R. B. Acquisition of motor skills; I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263-281.
2. AMMONS, R. B. Acquisition of motor skills: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393-411.
3. DENNY, M. R., FRISBY, N., & WEAVER, J. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48-54.
4. DOLLARD, J., & MILLER, N. E., *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
5. ERIKSEN, C. W. Psychological defences and "ego-strength" in the recall of completed and incomplete tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45-50.
6. ERIKSEN, C. W. Some personality correlates of stimulus generalization under stress. *J. abnorm. soc. Psychol.*, 1954, **49**, 562-566.
7. EYSENCK, H. J. *Dimensions of personality*. London: Kegan Paul, 1947.
8. EYSENCK, H. J. *The scientific study of personality*. London: Routledge and Kegan Paul, 1952.
9. EYSENCK, H. J. *The structure of human personality*. London: Methuen, 1953.
10. EYSENCK, H. J. *The psychology of politics*. London: Routledge & Kegan Paul, 1954.
11. EYSENCK, H. J. A dynamic theory of anxiety and hysteria. *J. ment. Sci.* 1955, **101**, 28-51.
12. EYSENCK, H. J. Cortical inhibition, figural after-effect, and the theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94-106.
13. EYSENCK, H. J. The inheritance of extraversion-introversion, *Acta Psychologica*, 1956, **XII**, 95-110.
14. EYSENCK, H. J., & PRESSL, D. B. The inheritance of neuroticism: an experimental study. *J. ment. Sci.*, 1951, **97**, 441-465.
15. FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1955, **2**, 143-156.
16. FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, in press.
17. FRANKS, C. M. An experimental study of conditioning as related to mental abnormality. Unpublished doctor's dissertation, Univer. of London, 1954.
18. HILDEBRAND, H. P. A factorial study of introversion-extraversion by means of objective tests. Unpublished doctor's dissertation, Univer. of London, 1953.
19. HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
20. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15-23.
21. KIMBLE, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500-510.
22. KIMBLE, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248-253.
23. KIMBLE, G. A., & BILODEAU, G. A. Work and rest as variables in cyclical motor learning. *J. exp. Psychol.*, 1949, **39**, 150-157.
24. KIMBLE, G. A., & HORENSTEIN, B. R. Reminiscence in motor learning as a function of length of interpolated rest. *J. exp. Psychol.*, 1948, **38**, 239-244.
25. KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355-359.
26. MCGEOCH, J. A., & IRION, A. L. *The psychology of human learning*. (2nd Rev. Ed.) London: Longmans Green, 1952.
27. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
28. NICHOLS, E. G. The relation between certain personality variables and the figural after-effect. Unpublished doctor's dissertation, Univer. of London, 1956.
29. PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.
30. SPENCE, F. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety.
31. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
32. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1952, **45**, 265-272.
33. TAYLOR, JANET A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81-92.
34. WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162-172.

Received December 19, 1955.

PERCEPTUAL ORGANIZATION AND INTELLIGENCE: A FURTHER STUDY¹

CLINTON DE SOTO² AND H. LEIBOWITZ

University of Wisconsin

SINCE the time of Galton (1), attempts have been made to measure intelligence with sensory discrimination tasks. These efforts have generally shown little success, and intelligence testing has consequently followed a different course. A new approach to the study of intelligence has been suggested by Krech and Calvin (2), who report a positive correlation between the ease with which *Ss* proceed through levels of perceptual organization and intelligence as measured by the Wechsler vocabulary test. (6). The purpose of the present study is to extend their experiments and to determine the effect of stimulus duration on this perceptual performance.

Krech and Calvin performed two experiments, with 28 college students as *Ss* in the first and 19 in the second. The apparatus for the perceptual task consisted of a Harvard tachistoscope, stimulus cards, a "reproduction board," and small black beads. The stimuli were patterns composed of black dots ($\frac{1}{16}$ inch squares) on a white background. The first experiment employed two patterns, an X figure and a rectangular array. The second experiment employed two rectangular arrays. The rectangular arrays included 72 dots linearly arranged in a 6×12 pattern (the entire pattern was approximately square). The dots were actually arranged in 6 rows of 12 dots per row, with $\frac{11}{64}$ inch between neighboring rows and $\frac{3}{64}$ inch between dots within rows. In the experiment using two rectangular arrays the rows ran vertically for one pattern ("columns") and horizontally for the other. The patterns were separated by $\frac{1}{2}$ inch. The reproduction board contained 900 concave hemispheres ($\frac{1}{2}$ -inch diameter) arranged closely in a 30×30 pattern.³

¹ This research was supported with funds provided by the Wisconsin State Legislature and by research grant M-1090 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

² Now at Johns Hopkins University.

³ Dr. Allen D. Calvin supplied us with a sample stimulus from which these specifications were estimated.

The *Ss* were told the test was of visual acuity and they were shown how to reproduce the stimulus by placing beads in the depressions on the board. They were told that they would repeatedly be presented briefly with two symbols, one of which was correct. They were to guess which was correct until they solved the problem and to reproduce both symbols after each presentation. The problem-solving demand was expected to elicit in each *S* the highest perceptual organization of which he was capable.

The stimuli were presented for 0.06-second exposures. The position order of the two patterns varied randomly over trials. The solution to the problem was a LRLR . . . alternation sequence. Trials continued until *S* achieved 6 consecutive correct choices or until 24 trials (16 trials in the second experiment) were completed.

Their results show that 28 *Ss* spaced the beads so as to reproduce rows and columns for some or all trials, while 19 *Ss* persistently reproduced the patterns as homogeneous patterns of equidistant dots. According to Krech and Calvin's theoretical discussion, the former (we will call them "high perceivers") showed a higher level of perceptual organization than the "low perceivers." In both experiments, high perceivers tended to achieve higher vocabulary scores than low perceivers (a mean of 33.6 as compared with 27.9 in the first experiment, and 30.6 as compared with 27.4 in the second experiment; these differences were significant as shown by the *t* test and a non-parametric test). A vocabulary score of 30.5 almost perfectly divided the high from the low perceivers. Further evidence that row and column reproduction is a high achievement is that few *Ss*, once attaining it, reverted to homogeneous reproductions.

Their theoretical discussion suggests that had the distance between the row groupings been greater or had the exposure time been longer, more of the *Ss* would have achieved the row grouping concept. In other words, they regard it as a threshold phenomenon.

The authors performed an experiment similar to Krech and Calvin's, varying stimulus exposure time and using somewhat modified experimental techniques.

METHOD

Thirty-six students from introductory psychology classes at the University of Wisconsin served as Ss. All had normal near vision as determined by a reading chart test.

A Harvard tachistoscope⁴ and stimulus patterns were employed which were essentially the same as in Krech and Calvin's second experiment with the exception that the two patterns of dots were separated by $1\frac{1}{8}$ inches rather than $\frac{1}{2}$ inch. A single center dot was presented in the adapting field of the tachistoscope and served as a fixation point prior to each stimulus exposure. The Ss observed binocularly at a viewing distance of 23 inches.

The Ss drew the stimulus by making dots with a number 2 pencil in the squares of graph paper uniformly ruled with light brown lines, 20 to an inch. This procedure was adopted to allow more accurate representations of the stimulus, particularly in size, than the reproduction board. It was assumed that "ability to draw" is scarcely more involved in placing dots in squares than in placing beads in holes.

We omitted the instruction to *S* to learn to choose the correct symbol and emphasized solely the importance of reproducing accurately what he had seen (calling it a "visual acuity" experiment, as did Krech and Calvin). The left-right position of the row and column patterns varied randomly over trials. The Ss drew both patterns on each trial.

Three exposure times were used: 0.01, 0.10, and 1.00 second. Illuminance level was constant at 10 foot-candles. Twelve Ss were assigned randomly to each exposure-time condition. Each *S* was given 16 trials at one of the three exposure times, then a 17th trial on which the stimulus was exposed for 30 seconds.

The Wechsler vocabulary test was administered prior to the perception test and the Ss were told that the two tests belonged to different research programs. The vocabulary scoring was done later for all Ss anonymously and simultaneously.

RESULTS

The over-all distribution of vocabulary score (mean = 29.17, $s = 3.23$) was similar to the distributions obtained by Krech and Calvin (mean = 31.98, $s = 4.29$; mean = 28.71, $s = 2.70$). The present sample is thus suitable in this respect for further study of their findings. The mean vocabulary score varied nonsignificantly among the exposure-time conditions.

The present results are similar to those of Krech and Calvin's in that the same propor-

⁴ Manufactured by Ralph Gerbrands, Arlington, Massachusetts.

TABLE 1
PERCEPTUAL PERFORMANCE AND VOCABULARY SCORE

Stimulus Exposure Time	High Perceivers	Low Perceivers	diff.	<i>t</i>
0.01 sec.				
<i>N</i>	9	3		
Mean Wechsler Vocabulary Score	28.56	26.17	2.39	1.09
0.10 sec.				
<i>N</i>	6	6		
Mean Wechsler Vocabulary Score	29.75	30.50	-0.75	.37
1.00 sec.				
<i>N</i>	7	5		
Mean Wechsler Vocabulary Score	29.00	30.00	-1.00	.57

tion of Ss, 60 per cent, were high perceivers (reproduced rows and columns accurately on at least one trial). We verified their finding that Ss tend to progress from homogeneous reproductions to row-column ones and rarely revert.

However, we found no relationship between perception and vocabulary scores, nor did we find a relationship between perception scores and stimulus-exposure time. Table 1 shows the number (*N*) of high and low perceivers in each exposure-time condition, together with their mean vocabulary score, the difference between the means of the high and low perceivers in each condition, and the *ts* for these differences. For each condition, a *t* of 2.23 would be required for significance at the .05 level, consequently none of the obtained *ts* were significant. We examined the operating characteristic function⁵ for *t* for the .05 confidence level, based upon 10 *df*, using a separate estimate (*s*) for each condition ($s = 3.29$, 3.50, and 2.98 for the 0.01-, 0.10-, and 1.00-second conditions, respectively.) The O-C function indicates that if the true difference between the mean vocabulary scores of the high and low perceivers had been 3.76 in the

⁵ Selection of the .05 significance level results in a 5 per cent chance of Type I errors (rejection of the null hypothesis H_0 when H_0 is true). In addition, use of the *t*-test, with 10 *df*, gives a certain chance of Type II errors (acceptance of H_0 when H_0 is false, i.e., failure to detect a true difference between the population means). The probability of Type II errors in case various absolute differences between means exist is given by the operating-characteristic (O-C) functions of the *t* test, using the sample estimates of σ . The O-C functions of a test are complements of its power functions. For some discussion of this see (5).

0.01-second condition, the probability of accepting H_0 (failing to reject it) would be .10. In other words, 90 per cent of repeated experiments would show a significant difference. If the true difference had been 3.99 in the 0.10-second condition, the probability of accepting H_0 would be .10. If the true difference had been 3.39 at 1.00 second, the probability of accepting H_0 would be .10. If the true difference had been 5.20 at 0.01 second, 5.53 at 0.10 second, and 4.70 at 1.00 second, the probability of accepting H_0 would be only .01 for each condition.

Because duration of exposure did not appear to affect the relative proportion of high and low perceivers (see the χ^2 test below), another t test was done between the high and low perceivers' mean vocabulary scores with the exposure groups combined. The high perceivers' mean was 29.02, and low perceivers' 29.39. The t for this difference is .33, far from significance. The O-C function for t for the .05% confidence level, $df = 34$, $s = 3.28$, shows that if the true difference were 1.87, the probability of accepting H_0 would be .10; and if the true difference were 2.53, the probability would be .01.

On Trial 17, with a 30-second exposure time, 28 Ss reproduced rows and columns and 8 did not. The mean vocabulary score of the high perceivers was 29.44; of the low perceivers, 29.09. This difference gives a t of .26, far from significance. The O-C function here (.05 level, $df = 34$, $s = 3.28$) is the same as that just described for all Ss on Trials 1-16 because s was almost identical. Again, if the true difference were 1.87, the probability of accepting H_0 would be .10; and if it were 2.53, the probability would be .01.

All these conjectural true differences, the first three fully independent, are smaller than the difference between means of 5.72 which Krech and Calvin found in their first experiment. They compare well with the difference of 3.15 obtained in their second experiment. Since the vocabulary distributions and the percentage of high perceivers were similar to theirs, it appears that each of our experimental conditions should have provided a fairly powerful test of the perception-intelligence hypothesis. However, our results were uniformly negative.

A possible relationship between time of ex-

posure and the percentage of high perceivers (or the intelligence cutoff value between high and low perceivers) was suggested by Krech and Calvin. Table 1 (the N entries) shows the number of high and low perceivers in each exposure time condition. The χ^2 for this table is 1.636, with $.30 < p < .50$, far from significance. Exposure time over the range from 0.01 to 1.00 second does not appear to affect perception of rows and columns. Such a finding is consistent with the previous observation that stimulus localization in the frontal plane is independent of the duration and luminance of the stimulus or stimuli (3, 4).

Only one of the 22 high perceivers on Trials 1-16 reverted to low perceiving on Trial 17, while 7 of the 14 low perceivers on Trials 1-16 became high perceivers on Trial 17. However, the correlated proportions χ^2 for these frequencies, corrected for continuity, is 3.13, not significant ($.05 < p < .10$).

The data were examined for sex differences (both males and females were used in all conditions) but none appeared.

American Council Psychological Examination (ACE) percentile ranks were obtained for most of the Ss. Kendall's rank-correlation coefficient Tau is .49 for the ACE total scores correlated with the vocabulary scores and .54 for the ACE L scores correlated with vocabulary. The ACE scores showed no relationship to perception scores.

Casual observations during the experiment suggest that many Ss who failed to reproduce rows and columns on the graph paper might have been credited with seeing them according to other measures. Some Ss, despite the instructions, attempted to draw dashes or solid lines following the stimulus rows and columns. When told to draw dots only, these Ss sometimes produced homogeneous dot patterns and received no credit for rows and columns. Other Ss spontaneously reported seeing rows and columns which they could not draw, attributing the rows and columns to factors other than spacing, such as the angle of light incidence and the angle of view. Many Ss were observed to draw the dots in horizontal rows (with lateral hand movements) for the row pattern and vertically for the column pattern, yet produce two identical homogeneous dot patterns. A few Ss, however, did report they had seen no difference or no internal difference between the patterns for some or all the trials.

DISCUSSION

The failure to repeat Krech and Calvin's findings demands careful study of the experimental techniques involved. The *S* population and method of stimulus presentation appear to be essentially identical. The instructions were somewhat different, those of the present study emphasizing accurate reproduction of the stimulus rather than learning the correct symbol. The percept-measuring responses were definitely different, although both would seem to involve the same type and level of critical abilities. Krech and Calvin's reproduction board allowed a maximum of 2 or 4 dots to an inch, while our graph paper allowed 10 or 20 to an inch. (The stimulus showed 9 dots to an inch within rows.) Perhaps this difference in the available sizes of the stimulus reproduction is important.

In any case, it would appear that the crucial factor or factors do not operate simply to make high perceptual organization easier or more difficult, but rather randomize it with respect to intelligence, and perhaps with respect to exposure time.

The facts that in the present experiment many low-vocabulary *Ss* did reproduce rows and columns, that 9 of 12 *Ss* did so even at the short exposure time of 0.01 second, and the indications that many *Ss* who did not reproduce them on graph paper were nevertheless perceiving them, suggest that almost all if not all *Ss* actually perceive them. If this is true, then any consistent failure of low-intelligence *Ss* to reproduce rows and columns may be a phenomenon more characteristic of the response than of the perceptual mechanism. The failure of 8 *Ss* to reproduce rows and columns even after a 30-second stimulus exposure bears out a response interpretation.

An alternative perceptual hypothesis may be suggested by the *Ss* who evidenced row-column percepts by reports and drawing techniques some trials before they actually reproduced rows and columns. Particularly striking are the 7 *Ss* who first reproduced them following the 30-second exposure even though many gave earlier evidence of seeing them. These *Ss* apparently could perceive rows and columns on brief exposure but could discover the spacing characteristics responsible only on a long exposure. Perhaps the perception of spacing determinants is a third level of perceptual

organization, following the homogeneous dots percept and a simple row-column percept.

Some elaboration of explanation such as those proposed seems required by the puzzling effects on measured perception of intelligence, stimulus duration, and response measures. The correlation of perception with intelligence (2), certainly a remarkable finding, appears to be revealed by the present study as complexly or obscurely determined and difficult to interpret.

SUMMARY

Krech and Calvin (2) obtained a significant positive correlation between intelligence and level of perceptual organization, which was measured by the ability to reproduce the rows and columns of patterns of dots viewed for 0.06 second. The present experiment, using modified techniques, showed no such correlation between intelligence and perception, although available evidence indicates the statistical tests possessed adequate power to test the hypothesis of such a correlation.

Krech and Calvin found that progression over trials was usually from low to high perceptual organization, rarely the reverse. Our data confirmed this finding. Contrary to hypothesis, the data show no correlation between perception and stimulus-exposure time from 0.01 to 1.00 second, and perhaps even through 30 seconds. There were indications in our data that many *Ss* who failed to demonstrate high perceptual organization with our stimulus-reproduction technique might have done so with other techniques. Possible interpretations of the data are discussed.

REFERENCES

1. GALTON, F. *Inquiries into human faculty and its development*. London: Macmillan, 1883.
2. KRECH, D., & CALVIN, A. Levels of perceptual organization and cognition. *J. abnorm. soc. Psychol.*, 1953, **48**, 394-400.
3. LEIBOWITZ, H., MYERS, N. A., & GRANT, D. A. Radial localization of a single stimulus as a function of luminance and duration of exposure. *J. opt. Soc. Amer.*, 1955, **45**, 76-78.
4. LEIBOWITZ, H. W., MYERS, N. A., & GRANT, D. A. Frequency of seeing and radial localization of single and multiple visual stimuli. *J. exp. Psychol.*, 1955, **50**, 369-373.
5. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 289-334.
6. WECHSLER, D. *The measurement of adult intelligence*. (2nd Ed.) Baltimore: Williams & Wilkins, 1941.

Received January 9, 1956.

IMPRESSIONS OF PERSONALITY, AUTHORITARIANISM, AND THE FAIT-ACCOMPLI EFFECT¹

I. H. PAUL

Austen Riggs Center, Stockbridge, Massachusetts

A UNITED STATES presidential election is an event of social impact and importance; a time when prominent political figures vie publicly for a position of conspicuous leadership, and people are sensitive to these figures as persons. This study is an investigation of people's impressions during the 1952 U. S. presidential election of the personalities of these leader-figures: how these impressions are affected both by the outcome of the election and by the people's orientation to authority.

Persons often face situations that, initially conflictful and equivocal, have become structured and settled by an accomplished fact beyond their control. Social psychologists have studied this situation during events such as elections and major legislative actions, and they repeatedly found that the *fait accompli* influences most people's opinions and attitudes and results in increased acceptance of the new state of affairs (3, 11). How general are the cognitive changes? Are impressions of the personalities of contending candidates in an election also affected by a *fait accompli*? The present study shows that they are.

Although the perception of another's personality is a crucial pivot in interpersonal dealings, it has only recently begun to attract the interest of psychologists. Pepitone has shown how a person who is seen to possess power tends, under certain conditions, to be viewed as approving, benevolent, and generally attractive (14). The valence of a person—his perceived attractiveness—seems also to be readily affected by external events (9), as well as by status and group membership (8), and it is well known that an important action of a person or one associated with him can radically change our impression of him.

¹ This article is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Pennsylvania. The writer wishes to thank Professor M. G. Preston who gave generous and invaluable supervision to every phase of this study. He also thanks Professor J. Wishner and A. Pepitone, members of his dissertation committee, for their guidance; and Miss M. Molitch for her assistance in collecting data.

Winning a U.S. presidential election is an important event: it results in a conspicuous change in status for the winner and endows him with power and success. Therefore, the first major expectation in this study is that the election outcome is followed by a change in people's impressions of the presidential candidates' personalities and results in an increase in the winner's attractiveness.

What role do a person's needs and dispositions play in such changes of impressions? Asch (2) recognized that the "needs and claims" which relate the perceiver to the perceived person may play a fundamental part in impressions of personality.² Such needs and claims, in the case of a person's relation to authority figures, are embodied in the concept of the "authoritarian personality." The conception—a broad and heuristically valuable one—has already stimulated extensive research.

Jones (10) recently reported an experiment that demonstrates the role authoritarianism plays in the formation of the person's first impressions of a leader-figure. Authoritarian persons tended to evaluate the leader more favorably than equalitarians did and, while in general they were less sensitive to the psychological or personality characteristics of others, they seemed more sensitive to such external variables as social status or position, and especially to power (defined as institutionally sanctioned authority).

The authoritarian person tends to conform and submit to authority (1, 12, 17). One of his submitting tactics might presumably consist of changing his impression of the person to whom he needs to submit. Such a tactic, in psychoanalytic terms, may be part of a rationalization or, in Festinger's terms (6), an attempt to bring his cognitions into *consonance* with a changed situation. A central assumption is that the type of change in impression which promotes such submission or consonance is, in

² A study by Preston, *et al.* (15) which shows that married persons' impressions of their spouses' personalities are a function of the marriage relationship itself is relevant to this point.

part, an increase in the favorability or attractiveness of that impression.

Therefore, the second major expectation in this study is that the authoritarian person in particular comes to view the winning candidate as an attractive person, i.e., he is more susceptible to the *fait-accompli* effect than is his equalitarian neighbor. A further expectation is that the authoritarian's alleged tendency to dichotomous thinking should result in greater differences between the ratings of his favored and his nonfavored leader-figures.

METHOD

Selection of groups. Three weeks before the election day in November of 1952 a short questionnaire-battery was administered to a pool of 320 volunteer business-school students of an introductory psychology course at the University of Pennsylvania. This battery consisted of: (a) The Short Authoritarian-Equalitarian Scale;³ (b) a statement of candidate-preference; (c) an index of preference-strength ("Indicate the strength of your feeling"); (d) an index of ego-involvement ("If your candidate does not win, how badly will you personally feel?"); (e) an open-ended question about the reasons for candidate-preference. A coding technique was used to maintain subject anonymity while facilitating the selection.

The results of this pretesting showed that 170 Ss favored Eisenhower and 150 Stevenson. There proved to be no relationship between A-E score and candidate-preference.

Four groups were selected from the upper and lower quartiles of the A-E distributions such that the two authoritarian groups—the one endorsing Eisenhower, the other Stevenson—had comparable A-E scores; and the same was true of the two equalitarian groups. Of 114 students asked to participate in the experiment, 96 appeared for the pre-election testing and, of these, 87 returned for the postelection testing. Seven Ss were later eliminated in order to equalize the four experimental groups at 20 Ss each.

The results of the pretesting showed that the groups differed in their preference-strength and ego-involvement. The proportion of authoritarian pro-Eisenhower Ss who professed *very strong* feelings was greater than the corresponding proportion of authoritarian pro-

Stevenson Ss ($CR = 2.11, p < .05$). The same was true for ego involvement. The differences between the equalitarian groups, though not as great, were consistent and in the opposite direction. These findings, of importance in their own right, are discussed below. However, the results on these items for the combined authoritarian groups did not differ from those for the combined equalitarian groups. The authoritarian Ss were somewhat more ego involved but this difference is not statistically significant.

Pre-election testing. In pre-election testing four to five days preceding the election day, the following battery was administered: (a) Impressions-of-Personality scales pertaining to Stevenson, Eisenhower, and S himself; (b) a repeat of the Short A-E Scale, and of the indices of candidate preference, preference strength, and ego involvement; (c) a number of biographical items such as age, socioeconomic status, and political orientation.⁴

Since the time of initial testing, no S had changed his candidate preference or his A-E classification, and the retest results for the items dealing with preference strength and ego involvement remained virtually identical.

Postelection testing. The postelection testing in which the Impressions-of-Personality scales were readministered took place two to three weeks following the election day. At the time of postelection testing, Eisenhower had only stepped into his new role but was not yet functioning in it, though he had made known two of his prospective cabinet choices—Dulles and Wilson. Stevenson ostensibly took his defeat with grace and good feelings.

Impressions-of-Personality scales. The work of Cattell (4) and of Fiske (7) on the factorial structures of personality ratings supplied a basis for devising the scale of impressions of personality. Twelve items were selected from Fiske's scale, 12 from Cattell's, and four new ones were added. The trait-items were classified into five a priori groups or classes, the first four being based upon the conclusions of Cattell and of Fiske: I Social Adaptability, II Emotional Control, III Conformity, IV Inquiring Intellect, and V Leadership Strength.⁵

For each candidate, S indicated his impression on each trait by placing a check mark in the appropriate position along an 8-point scale whose extremes are labeled and defined. To allow this rating to be judged with respect to favorability or attractiveness, S indicated, in a space beside each trait-item, at which pole he felt "*a President ought to fall*," i.e., which is the

³ The Short A-E Scale, devised and standardized by Sanford and Olders (18), is composed of items adapted from the California F scale plus other items, and was designed to meet the needs of a study of persons' general orientation to leadership. The scale was found to correlate .67 with the California F scale and the authors report a repeat reliability of .78 and split-half reliability of .58. In a subsequent study (17), a close relationship was demonstrated between a person's score on the scale and his overt orientation to leaders and to leadership phenomena. Recently, Eager and Smith (5) have reported data obtained in a children's summer-camp situation, which speak strongly for the predictive power of this scale.

⁴ The self-impressions and biographical items will not be considered in the present report.

⁵ Examples of each trait class: I, *Attentive to People* ("Interested in people, their troubles, their personalities. Makes friends with people and remembers their personal interests"); II, *Unshakable Poise* ("Self-possessed. Does not lose composure under emotional provocation"); III, *Independent-Minded* ("Thinks things out for himself. Adopts a clear and definite independent position"); IV, *Broad Interests* ("Has wide interests. Well informed on a wide variety of subjects. Shows intellectual curiosity"); V, *Persevering* ("Persists in any enterprise undertaken in spite of opposition. Steadfast").

TABLE 1

MEAN PRE- AND POSTELECTION RATINGS OF CANDIDATES BY AUTHORITARIANISM AND CANDIDATE PREFERENCE (20 CASES IN EACH SUBGROUP)

Trait-Class	Time	Authoritarian			Equalitarian			Total
		Pro-Eis.	Pro-Ste.	Total	Pro-Eis.	Pro-Ste.	Total	
Ratings of Eisenhower								
I. Social Adaptability	Pre-election	5.73	5.41	5.58	6.05	5.79	5.92	5.74
	Postelection	6.34	6.12	6.23	6.24	6.08	6.16	6.20
II. Emotional Control	Pre-election	6.08	5.10	5.59	6.40	5.88	6.14	5.87
	Postelection	6.56	6.07	6.32	6.66	6.37	6.52	6.42
III. Conformity	Pre-election	6.35	4.92	5.63	6.19	5.36	5.78	5.70
	Postelection	6.52	5.60	6.06	6.45	5.67	6.06	6.06
IV. Inquiring Intellect	Pre-election	6.22	4.50	5.38	5.95	5.03	5.49	5.44
	Postelection	6.53	5.58	6.06	6.26	5.70	5.98	6.02
V. Leadership Strength	Pre-election	6.21	5.32	5.77	6.09	5.53	5.81	5.79
	Postelection	6.47	6.10	6.29	6.34	5.91	6.14	6.21
Combined Trait-scale	Pre-election	6.12	5.05	5.58	6.13	5.52	5.82	5.70
	Postelection	6.49	5.90	6.19	6.39	5.94	6.17	6.18
Ratings of Stevenson								
I. Social Adaptability	Pre-election	4.98	6.06	5.53	5.65	6.29	5.97	5.75
	Postelection	5.42	6.36	5.89	5.83	6.20	6.02	5.95
II. Emotional Control	Pre-election	5.21	6.73	5.97	6.43	6.75	6.59	6.28
	Postelection	5.83	6.84	6.34	6.71	6.80	6.75	6.54
III. Conformity	Pre-election	5.13	6.41	5.77	5.58	6.75	6.17	5.97
	Postelection	5.45	6.33	5.89	6.04	6.58	6.31	6.10
IV. Inquiring Intellect	Pre-election	5.11	6.38	5.75	5.75	6.63	6.19	5.97
	Postelection	5.59	6.61	6.10	6.32	7.11	6.71	6.45
V. Leadership Strength	Pre-election	5.27	6.40	5.84	5.99	6.65	6.33	6.08
	Postelection	5.61	6.31	5.95	6.18	6.40	6.28	6.12
Combined Trait-scale	Pre-election	5.14	6.39	5.77	5.88	6.61	6.25	6.01
	Postelection	5.58	6.49	6.03	6.24	6.61	6.41	6.22

"desirable" rating. Each rating was then scored so that the 8 end represented the high-valued pole. The instructions, which were both printed and spoken, urged the Ss to reveal their impressions of the personality of the leader-figures rather than merely to give political opinions.

RESULTS

The ratings by the members of each experimental group were pooled to yield mean ratings for each candidate on each of the five classes of trait-items, as well as on the trait-scale as a whole. Table 1 presents these mean ratings in the pre-election and postelection tests.

The data were treated by analysis of variance.⁶ Since none of the quadruple-

⁶ The experimental design is a factorially confounded one: whereas each S rates both of the candidates in both testings, the Ss who compose the four experimental groups—the endorsement and authoritarian-equalitarian groups—are different. Therefore, higher-order-interaction variances involving combinations of Ss, as a variable, with these two variables cannot be extracted. Since the design furnishes no replication data, careful attention needs to be paid to the proper selection

interactions proved significant when tested against the quintuple-interaction, they were combined with it to furnish a composite highest-order-interaction variance.⁷ This variance then served as error-term to the triple-interactions. In testing each double-interaction and primary source the error-term selected was the next higher order interaction involving S or trait-classes (if that interaction itself proved significant). Table 2 presents the results of the analysis of variance.

1. Findings with respect to the expectation

of error-terms (higher-order-interactions) for the most valid assessment of statistical significance. It should be noted that the confounding eliminates some of the most sensitive error-terms that involve Ss as a variable, and interactions involving trait-classes have to be used in their place.

⁷ For a discussion of this procedure, see McNemar (13, p. 315). It should be noted that the quadruple-interaction involving S cannot be combined with the others, nor can it validly be tested against the quintuple-interaction which does not involve S. Therefore it remains as the highest-order-interaction to be used in assessing those sources in which S figures.

TABLE 2
ANALYSIS OF VARIANCE OF THE MEAN RATINGS

Source	df	Variance Estimate	Error Term	F Ratio	p
1. Of Eisenhower and Stevenson (L-F)	1	120.58	# 11	1.75	
2. Pre- and Postelection (I-F)	1	474.38	# 15	42.43	.001
3. Pro-Eisenhower & Pro-Stevenson (End)	1	22.13	# 34	2.77	
4. Authoritarian & Egalitarian (A-E)	1	290.25	# 34	36.46	.001
5. Trait-Classes (Tr)	4	69.05	# 19	9.93	.001
6. Subjects (S)	76	66.34	# 19	9.54	.001
7. L-F \times I-F	1	63.63	# 23	7.70	.01
8. L-F \times End	1	2258.26	# 25	54.32	.01
9. L-F \times A-E	1	104.17	# 34	13.09	.01
10. L-F \times Tr	4	36.52	# 27	6.43	.001
11. L-F \times S	76	68.93	# 27	12.13	.001
12. I-F \times End	1	1.09	# 29	3.11	
13. I-F \times A-E	1	31.59	# 34	3.97	<.05
14. I-F \times Tr	4	10.91	# 27	2.79	.05
15. I-F \times S	76	11.18	# 31	2.86	.01
16. End \times A-E	1	12.49	# 34	1.56	
17. End \times Tr	4	10.07	# 34	1.26	
18. A-E \times Tr	4	8.79	# 34	1.11	
19. Tr \times S	304	6.95	# 31	1.78	.01
20. L-F \times I-F \times End	1	109.39	# 34	13.74	.01
21. L-F \times I-F \times A-E	1	16.60	# 34	2.09	<.10
22. L-F \times I-F \times Tr	4	1.39	# 33	3.52	.01
23. L-F \times I-F \times S	76	8.27	# 33	20.67	.001
24. L-F \times End \times A-E	1	113.01	# 34	14.19	.01
25. L-F \times End \times Tr	4	41.57	# 34	5.22	.01
26. L-F \times A-E \times Tr	4	4.58	# 34	1.73	
27. L-F \times Tr \times S	304	5.68	# 33	14.27	.001
28. I-F \times End \times A-E	1	4.66	# 34	1.70	
29. I-F \times End \times Tr	4	3.69	# 34	2.16	
30. I-F \times A-E \times Tr	4	4.51	# 34	1.76	
31. I-F \times Tr \times S	304	3.91	# 33	9.33	.001
32. End \times A-E \times Tr	4	8.65	# 34	1.09	
33. L-F \times I-F \times Tr \times S	304	0.40			
34. Combined quadruple- and quintuple-interactions	21	7.96			

that people's impressions of the personalities of presidential candidates change following the election and this change involves an increase in the attractiveness of the winner:

Table 1 shows that for all Ss the increase in ratings on the combined trait-scale was greater in their impressions of Eisenhower than of Stevenson. While it is true that Stevenson, having been rated higher in the pre-election testing, continued afterwards to be rated very slightly higher (the difference is not statistically significant), the increment from the pre- to the postelection testing is significantly greater in the mean rating of Eisenhower.⁸

⁸ When dealing with changes along a linear scale, attention needs to be paid to the possible presence and influence of a scale-effect, because the poles of such a scale may exert a limiting influence on such change. In the case of the present study, since increase in rating is

The ratings of Eisenhower go from 5.70 to 6.18, an increase of .48, while those of Stevenson go from 6.01 to 6.22, an increase of .21 (the interaction L-F \times I-F, #6 in Table 2, represents this finding).

Inspection of Table 1 shows that in each of the five trait-classes the increase, for the experimental population as a whole, was greater in the impressions of Eisenhower than of Stevenson, though in the case of Inquiring Intellect (IV) it was virtually the same for both. In all the cases the pre-election rating of Eisenhower was the lower, but in two cases (Social Adaptivity and Strong Leadership) the postelection ratings of Eisenhower rose beyond those of Stevenson.

The differential increase in favor of Eisenhower varied with the trait-classes (L-F \times I-F \times Tr). It was greatest in the trait-class Strong Leadership (.38). The rank order of the remaining four is: Emotional Control (.29), Social Adaptability (.26), Conformity (.23), and Inquiring Intellect (.08).

The influence of the *fait accompli* was greater on those Ss who endorsed the loser (L-F \times I-F \times End). Table 1 shows that, while the ratings of Eisenhower increased more than those of Stevenson in the case of each group's rating regardless of the candidate initially favored, the relationship was stronger in the case of the unfavored candidate. In the latter case the ratings of Eisenhower on the combined trait-scale were initially lower but became slightly higher than those of Stevenson. Further examination of Table 1 reveals that ratings became less favorable in two instances (Conformity and Strong Leadership), both in ratings of Stevenson by those Ss who endorsed him. Moreover, this group showed a greater increase in its rating of Stevenson on Inquiring Intellect. The greatest increase in rating occurred for Eisenhower by pro-Stevenson Ss (.64); the next greatest for Stevenson by pro-Eisenhower Ss (.40); the

the preponderant finding and since the ratings are characteristically in the upper half of the scale, each finding needed to be scrutinized carefully to make sure it could not be due to a "ceiling-effect." It may be noted that the analysis of variance provides, in the interaction of I-F \times Ss, a partial index to scale-effect which can be used as the error-term where such an effect is suspected. Several other attacks on the problem were also made. Each of the major findings reported in this study has withstood rigorous scrutiny with regard to a "ceiling-effect."

next for Eisenhower by pro-Eisenhower Ss (.32); and the least for Stevenson by pro-Stevenson Ss (.04, which is not significantly different from zero: $t = 0.29$).

2. *Findings with respect to the expectation that the differential increase in favor of the winner is greater among authoritarian than among equalitarian persons.*

It has been seen that ratings of Eisenhower changed in a favorable direction to a greater extent than those of Stevenson. Such differential change in favor of the winner was more pronounced among authoritarian than among equalitarian Ss. The difference between these groups ($L-F \times I-F \times A-E$), however, does not attain statistical significance beyond the .10 level of confidence. Nevertheless, in the case of the authoritarians but not of the equalitarians, Eisenhower, initially lower, emerged finally higher than Stevenson. The differential increase in favor of Eisenhower was .35 among the authoritarians, and only .19 among the equalitarians.

Table 1 shows that in four of the five trait-classes the increment in the ratings of Eisenhower was greater than in that of Stevenson, both for the authoritarian and the equalitarian groups. The exception is the trait-class Inquiring Intellect where the increment of Stevenson was the greater for the equalitarian groups. A reversal of position (Eisenhower beginning as the lower of the two but emerging in the postelection as the higher) occurred only for one trait-class (Social Adaptability) among the equalitarian Ss, while it appeared in three trait-classes in the ratings of the authoritarian Ss (Social Adaptability, Conformity, and Strong Leadership). Change in an unfavorable direction appeared in only one instance, the rating of Stevenson by equalitarian Ss on Strong Leadership. While the rating of Stevenson on this trait-class by the authoritarian Ss increased very slightly, the difference is not statistically significant.

In the ratings of their nonfavored candidate the greater favorable change in impressions of Eisenhower was most pronounced for the authoritarian Ss. Among these Ss appeared the most conspicuous example of the *fait-accompl*i effect. Moreover, their initial ratings of Eisenhower were lower than of Stevensons and their final ratings higher in all trait-classes except one (Inquiring Intellect, where the final ratings are practically the same).

3. *Further findings.* Ratings of favored candidates were higher than ratings of nonfavored candidates ($L-F \times End$). Moreover, the discrepancies between Ss' ratings of favored and nonfavored leader-figures were greater in the pre-election testing than in the postelection testing ($L-F \times I-F \times End$). In both periods these discrepancies were greater for authoritarian Ss than for equalitarian Ss⁹—the mean pre-election discrepancy for authoritarian Ss being 1.16, for equalitarian Ss, 0.67; in the postelection testing, 0.75 and 0.45, respectively. Inspection of Table 1 suggests that this finding, which holds consistently for each of the trait-classes, is primarily due to the authoritarians' lower ratings of their nonfavored leader-figure. In each instance the authoritarian Ss rated their nonfavored candidate less favorably than did their equalitarian fellow-endorsers.

DISCUSSION

The results of this study bear out both of the major expectations. The *fait-accompl*i effect is reflected in the finding, demonstrated to some extent in each of the groups, that the ratings of Eisenhower's personality changed in the direction of favorability or attractiveness more than did the ratings of Stevenson. Furthermore, the extent of this differential change was greater among the authoritarian Ss than among the equalitarians.

The *fait-accompl*i effect was also greatest on those Ss, both authoritarian and equalitarian, who originally endorsed the loser. It stands to reason that this group should be most affected by the election outcome since each of its members was at variance with majority opinion and choice and had acquired a president whom he had viewed as a relatively unattractive person. The fact that his impression of the winner showed more improvement may be understood as an attempt to reach an equilibrium between his inclinations and the reality created by the election outcome—a "facilitative perception," a measure to attain "cognitive consonance," or a "rationalization."

Most of the Ss expressed awareness of the

⁹ However, the interaction which represents this finding ($L-F \times A-E \times End$) does not attain statistical significance. This may largely be due to the absence of a sufficiently sensitive higher-order-interaction variance which could serve as its error-term.

change in their impressions and many of them attributed it directly to the election outcome and its implications. Here are a few representative examples of their responses to the open-ended question concerning felt change:

I feel a little more confident in his ability as president. The change... is probably due only to the reason that he has been elected president. I can't explain myself too well, but I feel that Eisenhower is just a better man than I had imagined.

Personal feelings of Eisenhower have changed. Nothing that could be easily and categorically written down. I feel he is more of a man because of his victory.

The change in my impressions slightly toward Eisenhower is probably based on the respect to be shown to his office and the dire hope that he will turn out to be a better president than I thought he would have made.

Eisenhower being our hope for the next four years, becomes (out of desperation?) a better man.

I think I respect Ike more now, probably because: 1) he won by such a majority; 2) he is the president-elect; 3) he is picking fairly decent men for the top government posts.

The relationship between endorsement and favorableness of impression is clear-cut: positive endorsement goes with a more favorable view. However, in this connection some additional findings deserve consideration: namely, that in the post-election testing impressions of *both* leader-figures became more attractive, and those of *S's* initially nonendorsed candidate (whether winner or loser) did so to a greater extent.

Thomsen (19) found a shift in attitude intensity following the 1938 election from a U-shaped distribution toward a normal curve, which he interpreted in terms of a general decrease in intensity of attitude following the election, with a shift toward a more neutral position. It seems reasonable to assume, therefore, that the intensity of ego involvement with regard to candidate endorsement should also appreciably subside after the election. If we further assume that, in the tense and exciting atmosphere of the pre-election period, heightened ego involvement and attitude strength are influential primarily in bringing about a devaluation of a person's opposed candidate, then it follows that when this atmosphere is alleviated equilibrium may be re-established by improvement of this impression. This formulation, in addition to explaining the over-all improvement in the ratings of non-endorsed candidates, also leads to the important prediction that the impressions of the loser by those *Ss* who initially favor him,

should in fact not increase. We expect an increase for the winner by those who endorse him because of the various perceptual and motivational changes which are associated with the *fait-accompli* effect: he is a figure crowned with success and power. But in the case of the loser there is no reason to expect that those who initially endorse him should emerge with a more favorable impression. The data bear out this expectation. They show that it is precisely these latter impressions which showed no significant change.¹⁰

The role of ego involvement and attitude strength are further brought into prominence by the following findings. The selection procedure revealed that, although authoritarians and equalitarians were distributed equally in both endorsement groups, the authoritarian pro-Stevensonites felt less strongly and were less ego involved in their endorsement of Stevenson than were their equalitarian fellow-endorsers. Conversely, of the pro-Eisenhowerites it was the authoritarians who felt more strongly and were more ego involved. Now, in their impressions of the candidates it is the authoritarians, and especially those who initially endorsed Stevenson, who changed most in favor of Eisenhower. It appears, therefore, that strength of endorsement and extent of ego involvement are factors which were involved, at least partly, in the role which authoritarianism played in the *fait-accompli* effect. That these factors cannot alone account for the main findings is suggested by a number of considerations, among which is the observation that there is no consistent relationship in the results between the ratings by *Ss* and the extent of their ego involvement and strength of attitude.¹¹

The proposition concerning the authoritarians' tendency to think dichotomously has support in the finding that there was a greater discrepancy between their ratings of the candidates than between those of the equalitarians. Rokeach's (16) contention that ego

¹⁰ That these ratings fail to decrease may also need explanation. One may appeal, as Korchin does (11), to cultural factors in the U.S. which lend a loser some support and favor. Such values and mores are usually summed up in expressions like "on the side of the underdog," "sportsmanlike losing," and the like.

¹¹ For example, in the pre-election testing the authoritarian pro-Stevenson *Ss* rated Stevenson no differently than the authoritarian pro-Eisenhower *Ss* rated Eisenhower (see Table 1).

involvement is an important condition for the occurrence of this phenomenon is supported in the finding that this difference was greater in the pre-election period—a period which is most likely characterized by heightened ego involvement. These findings may be related to the authoritarians' devaluation of their non-favored candidate in the pre-election period since, following the election outcome, the decrease in disparity seems to have resulted primarily from an increase in these ratings.

SUMMARY

The results of this study permit the conclusion that, in a leadership contest such as a U.S. presidential election, a *fait-accompl*i effect occurs with respect to impressions of the personalities of the candidates. Furthermore, this effect is more pronounced among authoritarian persons than among equalitarians.

Ratings of the personalities of Eisenhower and Stevenson were obtained from authoritarian and equalitarian Ss one week before the 1952 election and again several weeks after the election. When these ratings are judged with respect to attractiveness or favorableness, the findings show a greater postelection change in a favorable direction for Eisenhower. This differential increase in favor of the winner was found for Ss who initially endorsed Eisenhower, for those who endorsed Stevenson, as well as for authoritarians and equalitarians. However, differential favorable change was greatest among the authoritarian Ss and among those who had endorsed the loser. The greatest differential increase occurred regularly for the trait-class Leadership Strength.

A further finding is that the discrepancy between a person's ratings of his favored and nonfavored candidates was greater before the election than after, and greater for the authoritarian person than for the equalitarian.

REFERENCES

1. ADORNO, W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
3. CANTRIL, H. *Gauging public opinion*. Princeton: Princeton Univer. Press, 1944.
4. CATTELL, R. B. The description of personality: basic traits resolved into clusters. *J. abnorm. soc. Psychol.*, 1943, **38**, 476-506.
5. EAGER, JOAN, & SMITH, M. B. A note on the validity of Sanford's authoritarian-equalitarian scale. *J. abnorm. soc. Psychol.*, 1952, **47**, 265-267.
6. FESTINGER, L. The relation between behavior and cognition. Unpublished paper read at Colo. Sympos. on Cognition, Boulder, 1955.
7. FISKE, D. W. Consistency of the factorial structures of personality ratings from different sources. *J. abnorm. soc. Psychol.*, 1949, **44**, 329-345.
8. HAIRE, M., & GRUNES, WILLA. F. Perceptual defenses: processes protecting an organized perception of another's personality. *Hum. Relat.*, 1950, **3**, 403-412.
9. HOROWITZ, M. W., LYONS, J., & PERLMUTTER, H. V. Induction of forces in discussion groups. *Hum. Relat.*, 1951, **4**, 57-75.
10. JONES, E. E. Authoritarianism as a determinant of first-impression formation. *J. Pers.*, 1954, **23**, 107-127.
11. KORCHIN, S. J. Restructuration of attitude following a national election: the *fait-accompl*i effect. *Amer. Psychologist*, 1948, **3**, 272. (Abstract)
12. MASLOW, A. H. The authoritarian personality. *J. soc. Psychol.*, 1943, **30**, 29-48.
13. McNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
14. PEPITONE, A. Motivational effects in social perception. *Hum. Relat.*, 1950, **3**, 57-77.
15. PRESTON, M. G., MUDD, E. H., PELTZ, W. L., & FROSCHER, H. B. Impressions of personality as a function of marital conflict. *J. abnorm. soc. Psychol.*, 1952, **47**, 326-337.
16. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259-278.
17. SANFORD, F. H. *Authoritarianism and leadership: a study of the follower's orientation to authority*. Philadelphia: Institute for Research in Human Relations, 1950.
18. SANFORD, F. H., & OLDER, H. *The short authoritarian-equalitarian scale*. Philadelphia: Institute for Research in Human Relations, 1950.
19. THOMSEN, A. What voters think of candidates before and after elections. *Publ. Opin. Quart.*, 1938, **2**, 269-274.

Received January 10, 1956.

RELATIONS OF TEMPERAMENT TO THE CHOICE OF VALUES

LYLE V. JONES AND CHARLES MORRIS

The University of Chicago

FIVE dimensions in the domain of value were isolated in a previous study, and empirical data from several nationality groups were interpreted in terms of that dimensional structure (2). The data were obtained from administration of the Morris *Ways to Live* schedule to male college students in India, Japan, China, Norway, and the United States. The *Ways to Live* schedule requires that a subject assign a rating to each of 13 paragraphs, each defining a "way to live." Each way to live represents a conception of desirable life, such as those found in prominent religious or ethical traditions. The document has been presented in its entirety by Morris and Jones (2).

In contrast with the earlier study, the present investigation focuses upon individual differences in expressed values among members of the same cultural group. In particular, it is hypothesized that the expressed value system of an individual is related to his temperament. For the purpose of the investigation, the value system is determined from response to the *Ways to Live* document, and scores for the areas appraised by the *Thurstone Temperament Schedule* (5) are employed as indices of temperament traits.

An important distinction should be emphasized between the value and the temperament domains, as here measured. The *Ways to Live* document assesses an individual's belief concerning the way he would like to live. In contrast, items of the temperament schedule refer to aspects of actual behavior. There is no necessary relationship between the two sets of responses. It remains an empirical issue to determine the pattern of relationships which may appear between temperament and values.

METHOD

Description of Variables

The Thirteen Ways to Live. Each subject is instructed to assign a numeral from one to seven to each of the paragraphs designating a way to live in accordance with the degree to which he would like to live that kind of life. These assigned numerals serve as scores for the

subject.¹ In Table 1 appear phrases serving to summarize the content of each of the *Ways to Live* paragraphs.

The Five Value Factors. Approximately 650 male college students of ages 20 or 21, from 21 colleges and universities throughout the United States, responded to the *Ways to Live* document during the period 1945 to 1954. Represented are both state and larger private universities. Within a school, students were chosen predominantly from general undergraduate courses in philosophy, the social sciences, and physical education, courses containing students with major interests in all parts of the university. The final sample of 250 was selected to be representative of this group.²

Product-moment correlation coefficients were computed among all pairs of ways as rated by these 250 persons; they appear in Table 2. A centroid factor analysis was performed and factors were rotated graphically to orthogonal simple structure (4, 6). In Table 3 is the centroid matrix, and the resulting rotated factor matrix is given in Table 4.³

The five value factors were then interpreted in terms of the factor solution of Table 4. A summary serving to identify the interpreted factors appears in Table 5. The thirteen variables in Table 1, together with the five variables of Table 5, define the value measures used in the present study.

The Seven Temperament Areas. In Table 6 are summarized the seven areas appraised by the *Thurstone Temperament Schedule*. From responses to the temperament schedule, a score may be determined for each subject on each of the seven temperament variables. The score is found by adding (a) the number of items in each temperament area for which response (yes or no) corresponds to the scoring key, and (b) one-half the number of items in that temperament area for which response is neither yes nor no.

Relating Values and Temperament

Both the *Ways to Live* document and the *Thurstone Temperament Schedule* were administered in 1951, 1953,

¹ For the purpose of obtaining product-moment correlations with other variables, it has been demonstrated that results from scores so defined do not differ substantially from results based upon the successive intervals scale values corresponding to the assigned numerals (cf. 2).

² Strictly speaking, the 250 were chosen from the larger group by a cyclic rather than a random sampling process. However, it is judged to be unlikely that bias was introduced.

³ Since this is a different sample from that analyzed earlier (2), which neither was selected from students throughout the United States nor was controlled for age, the factor loadings here reported differ somewhat from those presented in the earlier paper.

and 1954, to an additional 115 male undergraduate students from general classes at Harvard University, University of Chicago, Illinois Institute of Technology, and Illinois Wesleyan University. As a preliminary survey of relationships between value and temperament, product-moment correlation coefficients were calculated between the seven temperament variables and the thirteen ways to live.

In order to ascertain the relationship between the temperament areas and the value dimensions, a procedure (1) was employed for estimating the factor loadings of variables not included in the original analysis. The method allows for estimation of the loadings of the value factors upon each of the seven temperament variables. In geometric terms, these loadings represent the projections of a vector associated with each tem-

perament area upon the unit reference vectors representing the five orthogonal value factors.

RESULTS AND DISCUSSION

In Table 7 appear the correlations between ways of life and temperament areas. It is apparent that several columns (ways to live) and several rows (temperament areas) exhibit no significant correlation coefficients. Thus, for the values associated with ways to live numbers 1, 3, 4, and 10, there is no evidence of relationship with temperament. And for temperament areas A and E, there are observed no significant relationships with value.

The significant correlations of Table 7 may be organized so as to illustrate certain trends of relationship. Ways 5, 6, 8, and 12 exhibit similar patterns of correlations with the temperament variables. They tend to be related positively to temperament areas I, V, S, and D, and negatively to R. A second distinct cluster of ways to live consists of Ways 2, 7, 9, and 11, which tend to correlate negatively with temperament areas I, V, S, and D, and positively with R. A negative relationship between Way 13 and temperament variable R exhausts the significant correlations of Table 7.

The method of Dwyer (1, 3) was employed to estimate the loadings of the five value factors upon the temperament variables. It is instructive to consider the resulting matrix of estimated factor loadings (Table 8) as an addition of seven rows to the factor matrix of Table 4, and to review the definitions of the value factors in light of these added results.

Value Factor A, Social Restraint and Self

TABLE 1
SUMMARY OF THE 13 WAYS TO LIVE

Way	Summary of Content
1	Refinement, moderation, restraint; preservation of the best attainments of man.
2	Self-sufficiency, understanding of self; avoidance of outward activity.
3	Sympathy, concern for others; restraint of one's self-assertiveness.
4	Abandonment, sensuous enjoyment of life; solitude and sociality both are necessary.
5	Energetic, cooperative action for the purpose of group achievement and enjoyment.
6	Activity; constant striving for improved techniques to control nature and society.
7	Flexibility, diversity within self; accept something from all other paths of life.
8	Carefree, relaxed, secure enjoyment.
9	Quiet receptivity to nature yields a rich self.
10	Dignity, self-control; but no retreat from the world.
11	Give up the world and develop the inner self.
12	Outward, energetic activity; use of the body's energy.
13	Let oneself be used; remain close to persons and to nature.

TABLE 2
PRODUCT-MOMENT CORRELATION COEFFICIENTS AMONG THE 13 WAYS TO LIVE
(N = 250)

Test	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	044											
3	197	-061										
4	-216	122	-109									
5	-008	-266	097	047								
6	010	-085	017	-100	133							
7	-014	099	-083	075	-119	004						
8	012	-010	013	274	118	-075	077					
9	-104	221	145	-034	-055	-029	056	084				
10	265	059	130	-224	-010	067	-176	-097	-002			
11	-057	278	-024	118	-255	-189	098	-009	153	130		
12	029	-153	033	080	176	237	-074	001	-011	107	-078	
13	047	-034	177	-203	029	003	-154	-099	276	156	112	057

Note.—Decimal points are omitted in all entries.

TABLE 3
CENTROID FACTOR MATRIX

	I	II	III	IV	V	h^2
1	291	215	-.078	132	326	265
2	-.410	359	.078	259	157	397
3	250	208	136	-.155	162	177
4	-.416	-.327	286	.045	-.055	374
5	312	-.326	207	-.186	.066	290
6	291	-.185	.075	147	.098	159
7	-.293	-.051	-.083	.059	124	111
8	-.213	-.207	223	-.113	259	216
9	-.148	290	400	-.228	112	332
10	341	363	101	244	.012	313
11	-.343	370	158	183	-.162	336
12	309	-.233	356	254	-.067	346
13	237	394	270	-.245	-.149	363
Σa^2 13	093	083	047	035	025	(283)

Note.—Decimal points are omitted in all entries.

TABLE 4
ROTATED FACTOR MATRIX

	A	B	C	D	E	h^2
1	51	-.03	-.05	.02	-.02	264
2	10	-.28	.50	-.11	.20	392
3	25	.01	-.05	.34	-.03	181
4	-.44	12	.10	-.11	.38	374
5	-.04	.34	-.36	.20	.07	292
6	17	.30	-.18	-.06	.02	156
7	-.07	-.19	.02	-.18	.20	113
8	-.12	.02	-.08	.09	.44	223
9	-.03	-.09	.26	.47	.19	334
10	.41	.14	.22	.09	-.25	306
11	-.10	-.17	.54	-.02	-.02	331
12	.04	.58	.00	-.01	.01	339
13	.07	.02	.15	.51	-.27	361
Σa^2 13	059	056	066	055	046	(282)

Note.—Decimal points are omitted in all entries.

Control, contributes little to the temperament variables. The factor accounts for less than two per cent of the variance of the temperament scores ($\Sigma a^2/7 = .019$). The largest loading, $-.27$, appears on temperament variable V. A negative relationship here is reasonable, suggesting that the physically vigorous individual is unlikely to ascribe great virtue to social restraint or self control.

Factor B, represented by Ways 12, 5, and 6, defined as Enjoyment and Progress in Action, exhibits high estimated loadings on temperament variables Impulsive, Vigorous, Sociable, and Dominant. These results tend to reinforce the summary interpretation of Factor B

(Table 3), and suggest that those individuals who profess to value a purposive expenditure of energy also are those who do live vigorously, who are quick to make decisions, who are sociable, and who tend to exhibit initiative and qualities of leadership.

Factor C, Withdrawal and Self-Sufficiency, had been defined by Ways 2, 11, 9, and -5 . This factor now appears to be positively related to Reflective temperament, and exhibits smaller negative loadings on Sociable, Vigorous, and Impulsive temperament areas. Again, the behavioral correlates, from the temperament domain, seem consistent with the related expressions of value.

Value Factor D, Receptivity and Sympathetic Concern, exhibits only one moderately strong loading on the temperament variables, a negative loading on Reflective. It is reasonable that the reflective individual, who prefers theoretical thinking and who works alone, regards unfavorably a life devoted to the service of others.

Value Factor E, which was given the name of Self-Indulgence, exhibits near-zero loadings on all temperament areas, with the possible exception of Reflective (.21). That the factor contributes to the Reflective area probably could be attributed to the affinity of the reflective individual for Ways 2, 4, 7, and -13 (Table 7). A similar pattern is seen among the Factor E loadings upon these Ways (Table 4).

The communality estimates of Table 8 suggest that the Thurstone Temperament variables I, V, S, R, and D share a substantial portion of variance with the value domain, but that temperament variables E (Stable) and A (Active) are relatively independent of

TABLE 5
SUMMARY OF VALUE-FACTOR INTERPRETATIONS

Factor	Interpretation
A	Ways 1, 10, -4 , 3 Social restraint and self-control
B	Ways 12, 5, 6, -2 Enjoyment and progress in action
C	Ways 11, 2, -5 , 9 Withdrawal and self-sufficiency
D	Ways 13, 9, 3 Receptivity and sympathetic concern
E	Ways 8, 4, -13 , -10 Self-indulgence

TABLE 6
SUMMARY OF THE SEVEN THURSTONE TEMPERAMENT VARIABLES

Variable	Summary of Content
A: Active	Hurried in speech, walk, writing, driving and working; tends to be "on the go"; restless.
V: Vigorous	Physically active in sports and outdoor occupations; uses large muscle groups; tends toward great expenditure of energy.
I: Impulsive	Quick to make decisions; enjoys competition, easily changes from one task to another; tends to be daredevil, happy-go-lucky, carefree.
D: Dominant	Excels at public speaking, organizing social activities, persuading others; capable of taking initiative and responsibility; tends to be a leader.
E: Stable	Exhibits cheerful, even disposition; able to remain relaxed and calm in a crisis; does not fret about daily chores.
S: Sociable	Sympathetic, cooperative, receptive, and agreeable with others; makes friends easily and enjoys company of others.
R: Reflective	Prefers theoretical to practical problems; enjoys work requiring accuracy and detail; tends to be quiet, to work alone.

TABLE 7
PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN THURSTONE TEMPERAMENT VARIABLES AND WAYS TO LIVE
(N = 115)

Tempera- ment Variable	Way													
	1	2	3	4	5	6	7	8	9	10	11	12	13	
A	095	-102	-055	-108	152	114	-145	031	001	147	-073	096	084	Active
V	-005	-252†	-138	033	275	187	001	196	-275	007	-319	480	004	Vigorous
I	-034	-210*	-048	003	375	271	-279	282	-251	076	-247	395	-019	Impulsive
D	145	-191	-008	068	251	164	-224	089	-236	161	-124	187	-011	Dominant
E	-024	-032	-124	-124	003	010	037	093	-075	041	015	068	037	Stable
S	005	-336	050	065	400	172	-227	210	-200	169	-210	228	105	Sociable
R	-007	302	-118	126	-273	-027	184	-104	076	-024	246	-120	-299	Reflective

Note.—Decimal points are omitted in all entries.
* 184, Significant beyond .05
† 240, Significant beyond .01

TABLE 8
ESTIMATED VALUE FACTOR LOADINGS ON THURSTONE TEMPERAMENT VARIABLES

Tempera- ment Variable	Value Factor					
	A	B	C	D	E	h ²
A	15	24	-06	09	-07	097
V	-27	69	-26	-13	-06	637
I	07	75	-24	-04	13	644
D	17	48	-15	-08	00	288
E	01	06	00	-06	-06	011
S	-02	55	-33	17	00	441
R	06	-20	38	-41	21	400
$\sum a^2$						
7	019	239	058	033	010	(359)

Note.—Decimal points are omitted in all entries.

value, as measured by the *Ways to Live* document.
The final row of Table 8 indicates the estimated proportion of variance of the temper-

ament variables which could be accounted for by each of the value factors. Factor B contributes strongly to the temperament domain. Factors C and D contribute less strongly, while Factors A and E are relatively independent of temperament.
The methods applied in this study allow the possibility of further rotation of the value factors to achieve more satisfactory simple structure, based upon the augmented factor matrix (Table 8 and Table 4). The graphical plots of each of the five factors against each other, based upon all twenty variables, do not suggest that the structure would be improved with additional rotations. Thus, the factor solution based upon only the thirteen ways of life is not modified by the addition of factor loadings upon the seven temperament variables.
The results of this study strongly suggest substantial relationships between the domain of temperament and that of value, especially

for values represented by Factors B and C, Enjoyment and Progress in Action, and Withdrawal and Self-Sufficiency. Relations are less marked between temperament areas and value Factor D, Receptivity, and are much less marked between temperament areas and value Factors, A and E, Social Restraint, and Self-Indulgence.

It does not follow, of course, that no temperament or personality traits are operative in the favor shown by individuals to the value Factors A and E. Temperament traits have, in this study, been measured only by the *Thurstone Temperament Schedule*; it may well be that other instruments would reveal further relations to the values isolated by *Ways to Live*, or to values as isolated by other instruments. Even if such further relations were found, however, their discovery would not indicate that temperamental traits are the sole determinants of value choices. Biological differences among individuals may serve as determinants; value patterns dominant in the culture in which choices are made may influence such choices; the particular social roles of an individual may contribute their influence.⁴

⁴ A study of biological and social determinants of value is included in the book *Varieties of Human Value*,

The limited purpose of the present study was to explore the nature of relationships between temperamental traits and the selection of values, and to illustrate the availability of factor-analytic methods for the study of such relations between temperamental and valuational domains.

REFERENCES

1. DWYER, P. S. The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika*, 1937, **2**, 173-178.
2. MORRIS, C., & JONES, L. V. Value scales and dimensions. *J. abnorm. soc. Psychol.*, 1955, **51**, 523-535.
3. MOSIER, C. I. A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika*, 1938, **3**, 297-299.
4. THURSTONE, L. L. *Multiple-factor analysis*. Chicago: Univer. of Chicago Press, 1947.
5. THURSTONE, L. L. *The Thurstone Temperament Schedule*. Chicago: Science Research Associates, 1950.
6. ZIMMERMAN, W. S. A simple graphical method for orthogonal rotation of axes. *Psychometrika*, 1946, **11**, 51-56.

Received January 12, 1956.

by Charles Morris, published in 1956 by the University of Chicago Press. Some use is also made in this book of Cattell's *16 Personality Factor Test* and Allport and Vernon's *Study of Values*.

ENHANCEMENT OF CONTRAST AS COMPOSITE HABIT¹

DONALD T. CAMPBELL

Northwestern University

ONE important aspect of the general syndrome of social stereotyping is *enhancement of contrast* or the exaggeration of relative differences between social groups. Thus if on intelligence tests in New York city schools Jewish students test slightly higher than white Christian students, and Negro students slightly lower, these small differences are exaggerated in social stereotypes about the students into the judgments that Negroes are "dumb" and Jews are "smart." The perceived direction of difference is in agreement with the facts, but the proportional difference is exaggerated, the contrast enhanced, and, of course, unwarranted genetic causation imputed.

Objective studies demonstrating the effect are few in number, for such demonstration requires social judgments in situations where the appropriate facts are known, and in which a language of social judgment commensurate with the social facts can be employed. A slight degree of such bias has recently been demonstrated for white students' estimates of the examination scores of their Negro and white classmates (3). Kephart (5) has demonstrated a quite marked tendency for policemen to exaggerate the proportion of the crimes in their own districts committed by Negroes, and by inference this implies an exaggeration of the Negro-white differential in crime rates. The other available data are somewhat tangential to the usual discussions of stereotyping. In the Hartshorne and May (4) studies of character, it was possible to compare boys and girls on both objective tests of character and on reputed character as judged by classmates. They suggest "prejudice" of the form here called "enhancement of contrast" as the explanation of instances in which objective tests showed slight sex differences and reputation

tests showed large ones. Thus on objective tests of "service," girls scored slightly higher than did boys (*CR* 1.9), while in reputation the difference was in the same direction but was much larger (*CR* 7.9). For "persistence" the same bias in reputational judgments was obtained (*CR*s 1.7 and 7.6). But for "inhibition" the sex difference appeared equally in objective tests and reputation (*CR*s 5.5 and 5.0). The comparison was not made for measures of "deceit."

While the empirical evidence is limited, it may be judged in sufficient accord with many less formal observations to establish enhancement of contrast as a frequent characteristic of social judgments. In specific instances this bias might be attributed to the effects of traditional attitudes transmitted from generation to generation—a learned mistaken generalization, biasing judgments of specific individuals in the concrete situation. Although this explanation remains unrefuted for the data cited, it is the effort of this paper to derive such a tendency toward bias from more general cognitive principles that would generate it even in the absence of biased traditional attitudes.

Enhancement of contrast seems to be anticipated by several approaches to cognitive functioning. In the philosophical criticism of common thought processes it appears under such labels as the tendency toward thinking in all-or-none, black-or-white, terms. Among the stubbornly recurrent errors of inference found in the teaching of logic is the error of the undistributed middle term. This error can be interpreted psychologically as a tendency to treat the pairs of terms as either totally overlapping or totally different, failing to note the possibility of partial overlap. The phraseology itself comes from the language of perception, and is related to the classification of optical illusions into errors of contrast and errors of assimilation. The view is compatible with the general spirit of Gestalt psychology, although a specific derivation might be difficult to achieve.

In calling attention to this bias, or to any other aspect of stereotyping, the emphasis of

¹ This study was initiated while the author was a visiting associate professor at Yale University, Spring, 1955. It has been supported by the Yale program of research on attitudes and communication which is financed by a grant from the Rockefeller Foundation and which is under the direction of Carl I. Hovland. Lyman W. Porter, Sarah Kendrick Martin, and Curtis E. Thomsen have contributed to the design and execution of the study.

the social psychologist is not only upon the fact of cognitive imperfection, but upon the *kind* of imperfection. The emphasis is on the constant or systematic errors. Learning theory, with its primary focus on the *degree* of perfection of the cognitive achievement rather than on the *kind* of imperfection, might seem less likely to be related to the principle. Yet there are treatments of systematic error to be found in the learning literature. A most pervasive principle predicting biased responses is that of stimulus generalization or stimulus confusion on the basis of stimulus similarity. Applied to the judgment of individuals in situations of group difference, the principle of stimulus generalization seems to predict an inevitable enhancement of contrast at stages of imperfect learning.

Consider as a paired associates task the problem of a student learning about the classroom competence of each of his Negro and white classmates, in a situation where on the average the white students do better. If skin color provides a significant degree of stimulus similarity, then stimulus generalization will lead to the confusion of one white student with other white students, one Negro student with other Negro students, prior to the completion of learning. The net result will be a tendency to confuse a given white student with the other white students, and thus to distort the judgments of individual white students in the direction of the average white student, and correspondingly for the Negro students. This distortion will be greatest for those students deviating most from their similarity group, including the least competent whites and the most competent Negroes, resulting in an underestimation of the intergroup overlap. (A more detailed derivation is provided below, in terms of the experimental analogue.)

This derivation seems to place the enhancement of contrast firmly within the range of phenomena covered by learning theory. However, it involves an extension of theory into a setting not yet studied. The interest of the social psychologist is often in the cumulative effects of a large number of "habits" or "associations" achieved contemporaneously, and not carried to perfection. This is a problem of multiple habit combination, focused not on the degree of achievement of the elementary association units, but rather on the collective resultant response tendencies. The term *com-*

posite habit is used to refer to this problem. In our hypothetical classroom setting there is interest in the composite habit resulting from the imperfect learning of a number of paired associations of students and classroom performances. Because of the remoteness of the derivation from any actual learning experiment, it seemed desirable to check it through an appropriate laboratory task. The study reported below was conducted for this purpose.

METHOD

The apparatus consisted of a board $8\frac{1}{2} \times 26$ in. with raised edges to hold in place a sliding aluminum cover. There were two apertures in this cover, in horizontal alignment with each other. The smaller one (1×1.5 cm.) at the left exposed a nonsense syllable. To the right of this was a long rectangular opening (1×12.5 cm.) through which only blank white paper could be seen. By searching with his pencil in this rectangular area, the *S* could find a vertical slot, cut in an underneath cardboard key, but not visible because covered with a record sheet, a carbon, and the top piece of blank paper. The *S*'s task was to learn the correct location of the slot for each syllable. The *S* made an initial guess (which he circled to make the record clear) and then starting from the initial guess searched the area for the slot in such a fashion as to minimize the length of the search path. When the slot was found, he moved the aluminum cover one step, exposing the next syllable and search space, and so on through the 20 items on the stencil. At this point the *E* quickly removed the stencil sheet assembly and inserted a new one, containing the same 20 items in a different order. The stencil sheet and the covering sheets were stapled together to insure easy and accurate placement. Eight random item arrangements were used, and these repeated in random orders until learning reached the criterion.

In selecting the groups of syllables there was an effort to provide two definite groups in terms of similarity, without the bases of grouping being obvious to the *Ss*. One similarity group consisted of seven syllables all having a middle *E* (QEB, GEP, ZER, BEH, FEQ, KEV, PEZ); the other consisted of seven syllables all having a terminal *X* (CUX, DAX, ZOX, BIX, GUX, KAX, RIX). To further reduce the likelihood that the *Ss* perceive the classification and thus treat the assignment as a task for concept formation or problem solving, there was a third group of syllables having low similarity both to one another and to either of the main groups (FUB, VAF, QIW, RUC, FOJ, LIH). As far as could be ascertained, none of the *Ss* discovered the similarity groupings or the purpose of the experiment.

Centered within the 12.5 cm. search space a distance of 5.5 cm. was selected for the placement of the vertical slots. This distance was divided into 11 loci, .5 cm. apart, and numbered for designation purposes from left to right. The seven —E— syllables were associated with positions 1 through 7, in the order listed above. The seven —X syllables were associated respectively with positions 5 through 11. The 6 buffer syllables were assigned positions 1, 3, 5, 7, 9, and 11, respectively. The

numbering and spacing of the loci were not visible to the Ss, and the smooth surface of the aluminum cover provided no specific position cues.

The Ss, who were paid, were recruited primarily from introductory psychology courses. The first group of eight worked in two-hour sessions, up to a total of eight hours, or until the learning criterion was reached. The next twelve worked one hour at a time for up to five hours. The criterion at which learning was stopped was an average of one spacing unit (.5 cm.) of error per item for five pages or stencils. The speed of working and rate of learning varied widely. The Ss took from 1 to 6 hours to reach the criterion, and from 15 to 115 trials, or stencils.

THE PREDICTION

In a nonsense syllable experiment, the troublesome problem of defining stimulus similarity by stimulus characteristics can be solved in part by defining similarity in terms of identical elements. To provide a detailed illustration the character of the prediction from stimulus generalization, the following quantitative assumptions have been made: Each shared letter shall provide 1 unit of similarity; a shared letter in the same position in the syllable 2 units; a common two-letter sequence in the same position in the syllable 2 additional units. Under this scoring the "similarity" of a syllable with itself is 10 units. The similarity with another syllable from its similarity group is at least 2 units. Table 1 has been prepared on this basis, showing for each syllable

the similarity units associating it with each one of the response positions, after at least one trial. Note the irregularities introduced through extraneous incidental similarities, including similarities with the buffer syllables.

This distribution of values can be taken as indicating the tendency of one syllable to be confused with the various others, or to be responded to as though it were one of the others. It can be taken as a plot of the response strengths (elicited through stimulus similarity) for each one of the response positions. The average resultant response to a given syllable is assumed to be some kind of an average of the competing response tendencies. Simple means of the distributions have been computed and these are presented graphically in Fig. 1. This averaging can be envisaged as occurring in two different ways, although the present analysis does not discriminate between them. On the one hand, each individual response might represent a spatial compromise. On the other hand, over a population of trials and persons, each position may occur in proportion to its strength, thus giving an average of responses the compromise character. The assumption of response combination or compromise is probably not useful for many learning situations, but does seem appropriate for that class of response called "judgments."

TABLE 1
PREDICTED DISTRIBUTION OF STRENGTHS OF EACH POSITIONAL RESPONSE FOR EACH SYLLABLE ON THE BASIS OF
STIMULUS GENERALIZATION STATED IN TERMS OF IDENTICAL ELEMENTS

		Response Positions											Mean
		1	2	3	4	5	6	7	8	9	10	11	
True Positions	Buffer Syllables	FUB		VAF		QIW		RUC		FOJ		LIH	
	—E—'s	QEB	GEP	ZER	BEH	FEQ	KEV	PEZ					
	—X's					CUX	DAX	ZOX	BIX	GUX	KAX	RIX	
Response Strengths for:													
QEB		12	2	2	3	5	2	2	1				3.21
GEP		2	10	2	2	2	2	3		2			3.88
ZER		2	2	10	2	2	2	5				1	4.41
BEH		4	2	2	10	2	2	2	2			2	4.57
FEQ		5	2	3	2	11	2	2		2			4.31
KEV		2	2	3	2	2	10	2			2		5.08
PEZ		2	3	3	2	2	2	11			2		4.96
CUX		2				10	2	5	2	6	2	2	6.81
DAX				2		2	10	2	2	2	6	2	7.36
ZOX				2		2	2	11	2	4	2	2	7.37
BIX		2			2	4	2	2	10	2	2	8	7.71
GUX		2	2			6	2	4	2	10	2	2	7.00
KAX				2		2	8	2	2	2	10	2	7.80
RIX				1		4	2	4	6	2	2	12	8.52

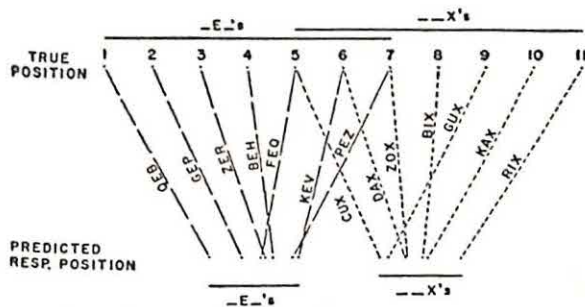


FIG. 1. PREDICTED MEAN RESPONSE POSITIONS

In Fig. 1 can be seen the predicted distortion of the extreme members of each similarity group toward the group mean, and the resulting reduction of overlap between the two groups.

RESULTS

Each trial, or sheet of 20 items, was scored in two ways, paying attention only to the initial responses to each item. For the degree of learning, the number of .5 cm. units of error for each of the 20 items was summated. For a measure of enhancement of contrast, attention was paid only to the syllables of the overlap region, positions 5, 6, 7, syllables FEQ, KEV, and PEZ; CUX, DAX, and ZOX. The sum of the initial guesses, scored in terms of the basic eleven steps appropriately extended, for the three E syllables was subtracted from the sum of the initial guesses for the three X syllables. In terms of the correct positions, the two sets of 3 syllables were equivalent. Random errors around the correct positions would result in differences with a mean of zero. The expected bias would distort E syllables to the left, X syllables to the right, resulting in a positive difference score. This bias, being based on generalization of learned associations, was expected only after some degree of learning had taken place, and of course would disappear as learning approached perfection.

The first trial error scores ranged from 74 to 177, and some Ss showed initial increases in errors and long plateaus before any improvement. An error score of 60 was selected as a suitable point to take as evidence of some learning. All bias scores were summated from that point to the point where the error score averaged 20 or below. Such mid-learning bias scores were computed on all 20 Ss. For 19 the sum was positive, indicating a net error in the

enhancement of contrast direction. This outcome provides a one-tailed p value of $<.00002$, confirming the prediction with a high degree of significance. While the bias could not be expected to show before some learning has taken place, it is hard to designate what constitutes a "chance" score on this particular task, and hence what indicates the beginning of improvement. For this reason, the selection of the initial point of 60 error units per trial may seem somewhat arbitrary. If bias scores are summed for all trials, including the first one, up to the criterion for learning, the score is positive for the 16 out of 20 Ss, significant at the $<.006$ level. The few additional exceptions that occur seem explainable in terms of random but large "mistaken notions" about certain syllables in the prelearning plateau.

Figure 2 shows the average trend in the bias score over the course of learning. This is a Vincent-type curve, achieved by computing for each S separately the average score for five equal-trial steps between his 60 and 20 error points. These five values from each S were then averaged. Figure 2 confirms the expectation that the bias should first rise as learning starts and then fall as learning approaches perfection.

Figure 3 shows for the second stage of learn-

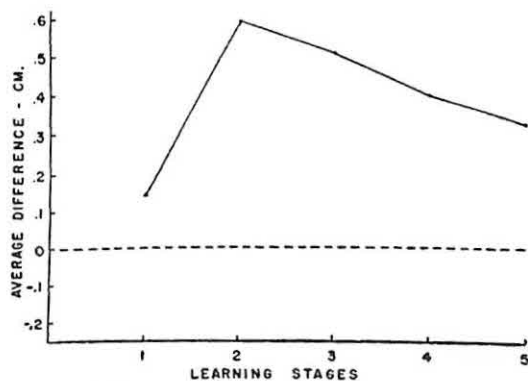


FIG. 2. DEGREE OF CONTRAST ENHANCEMENT BY STAGE OF LEARNING

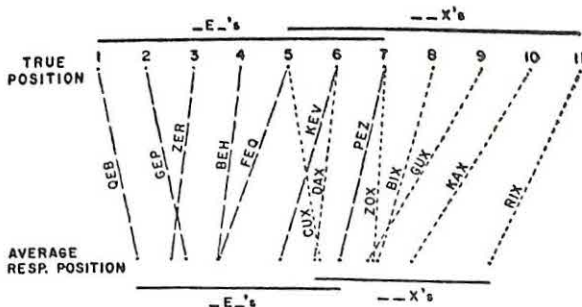


FIG. 3. AVERAGE RESPONSE POSITIONS AT LEARNING STAGE 2

ing the average response position for each of the syllables in the two similarity groups, and thus gives a graphic, over-all picture of the "stereotyping." In an experiment such as this, there is a problem of sampling syllables. The detailed interpretation of the present data is weakened by the fact that only this one set of syllables was used, and that a specific syllable was assigned to the same position for all Ss. Differences in the rates of learning specific to each syllable could affect the apparent distortion shown. With these qualifications in mind, Fig. 3 can be examined for general conformity to further implications of the stimulus generalization principle. Note in Fig. 1 that stimulus generalization *does not* predict that the *absolute* mean difference between the two groups as a whole will be exaggerated. Rather, members of each group are drawn toward their own group mean. And since some similarity is shared between the two groups, the group means will tend to be drawn together. These predictions seem roughly confirmed. Further, not only will the individuals in the overlap positions be distorted, but also those which deviate from their group mean at the non-overlap extremes. Because of the similarity between the two groups, these outward extremes should be even more distorted than the overlapping extremes. This prediction seems roughly confirmed, although obscured by a general leftward bias. These predictions and findings correspond in part to the social-psychological notion that stereotyping involves both reduction of overlap between groups and underestimation of variation within groups.

In this setting, reduction of overlap is not synonymous with exaggeration of the average absolute difference between groups. It is rather the exaggeration of the relative difference—relative to the over-all range of judgments made. This point might be a source of confusion in comparing the present predictions and findings with the study of children's estimates of classmates' examination grades (3). In that study, the results were reported as an exaggeration of differences between groups, without making a distinction between relative and absolute differences. Relative differences were at issue, however, as the judgments and test scores were independently converted into standard scores in terms of their own distributions. Were the distribution of average re-

sponse positions in Fig. 3 to be transposed into standard scores and the mean difference between the —E—'s and the —X's to be computed in terms of these standard scores, a comparable exaggeration of group differences would be found. A similar treatment is provided in the Hartshorne and May data cited (4), through the use of critical ratios to measure overlap. In natural situations, it will be very difficult to use absolute comparisons with confidence, and this will limit the kind of predictions which can be tested. For example, it would be impossible with standard scores to test for the underestimation of differences within groups separately from the exaggeration of differences between groups.

DISCUSSION

The derivation of the bias of enhancement of contrast from the principle of stimulus generalization seems justified as an extension of the facts of learning and by the illustrative experiment here reported. This approach makes the bias a natural, automatic, and inevitable aspect of imperfect learning about the individual members of overlapping groups. To say this does not remove the onus of bias nor correct the injustices that derive from it.

The derivation, like others based upon cognitive processes, leads to the prediction of the enhancement of contrast on the part of all observers of the situation, including the Negro students as well as the white ones in the illustration we have employed. It must be noted that this expectation is not confirmed in the one relevant study available (3), where the Negro-white differential in exam grades was exaggerated by the white students but not by the Negro students. The total cumulative research lying behind the stimulus generalization principle is so great that one must expect the principle to be operating in this situation, although possibly in conjunction with other sources of distortion which counteract it. More extensive data are needed on this point.

At the present time in social psychology the products of experience are termed *perceptions* rather than *habits* or *response tendencies*, particularly where the products are imperfect or biased. And while introductory psychology, experimental psychology, and learning theory have become more and more dominated by the

behavioristic stimulus-response language, social and personality psychology have become more dominated by subjective, perceptual, phenomenological, or conscious-experience terminologies. Upon careful inspection it often turns out that these languages are largely interchangeable, are expressing the same kinds of environmental-behavioral regularities, are making the same kinds of predictions. But this interchangeability is so little noted that the semantic confusion remains a big problem for an integrated psychology. In the initial discussion we started out with "judgments of" persons. These are usually interpreted as "perceptions of" persons, but our main discussion has been in terms of "responses to" persons. While this led to a different choice of theoretical models, the phenomena covered have been no different. Thus although the lower line in Fig. 3 has been labeled "average response position," it could have been labeled "perceived position" with equal appropriateness. In dealing with the confusion of stimuli or distortion through stimulus generalization, the behaviorist tends to use language implying the distortion of response to an unchanging stimulus. The subjective perceptionist tends to use language implying distortion in the perceived object. While in specific instances like the present one the intersubstitutability of languages may seem obvious, the superficial dissimilarities are often so great that well-established learning phenomena may seem new when rediscovered in perceptual terms (2, pp. 167, 168). Were all perceptionists as straightforward as Asch (1, pp. 158-59) in positing an isomorphism of subjective experience and action (and even he has done so explicitly for but a limited range of conditions) the identity of the formulations would become

very patent indeed. Focusing attention in learning studies on the *actual resultant responses* rather than upon the degree of perfection of responses may help make the isomorphism of perception and response more obvious.

SUMMARY

Enhancement of contrast is noted as a bias occurring in judgments made of individuals belonging to two overlapping groups. This bias is derived from the learning theory principle of stimulus generalization as a function of similarity. An experimental analogue is tested, in which two similarity groups of nonsense syllables are associated with a partially overlapping series of loci on a spatial continuum. At stages of incomplete learning the predicted enhancement of contrast or exaggeration of difference between the two similarity groups of syllables is found. The principle is discussed in terms of social stereotypes, and as an illustration of the equivalence of perceptual and behavioral languages for describing the results of experience.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
2. CAMPBELL, D. T. Operational delineation of "What is learned" via the transposition experiment. *Psychol. Rev.*, 1954, **61**, 167-174.
3. CLARKE, R. B., & CAMPBELL, D. T. A demonstration of bias in estimates of Negro ability. *J. abnorm. soc. Psychol.*, 1955, **51**, 585-588.
4. HARTSHORNE, H., & MAY, M. A. *Studies in the nature of character: I. Studies in deceit. II. Studies in service and self-control*. New York: Macmillan, 1928, 1929.
5. KEPHART, W. M. Negro visibility. *Amer. sociol. Rev.*, 1954, **19**, 462-467.

Received January 18, 1956.

A FACTORIAL STUDY OF DOGMATISM AND RELATED CONCEPTS

MILTON ROKEACH¹

Michigan State University

AND

BENJAMIN FRUCHTER

The University of Texas

THE major purpose of this study is to determine whether the concept of dogmatism (6), and its operational measurement by means of the Dogmatism scale (8), has content, as determined from its analysis in a battery with related scales, which distinguishes it from other well-known concepts such as authoritarianism, ethnocentrism, and rigidity. Elsewhere the concept of dogmatism has been given the following threefold definition: it is "(a) a relatively closed cognitive organization of beliefs and disbeliefs about reality, (b) organized around a central set of beliefs about absolute authority which, in turn, (c) provides a framework for patterns of intolerance and qualified tolerance toward others" (6, p. 195). In line with this definition, the Dogmatism scale is conceived to measure both *general* authoritarianism and *general* intolerance which is not restricted to any specific ideological content. In contrast, the F scale (1) is assumed to measure primarily right authoritarianism rather than general authoritarianism and the E scale (1) to measure rightist forms of intolerance rather than general intolerance.

From dogmatism thus defined follows a distinction between dogmatism and rigidity. While dogmatism refers to total *systems* of beliefs and disbeliefs which are closed or resist change, rigidity refers only to single beliefs, sets, or expectancies which resist change. This distinction has received some experimental support in a study by Rokeach, McGovney, and Denny (7). In that study subjects were presented with a cognitive task the solution to which required the overcoming of three separate sets and then their integration. Subjects scoring high on the Gough-Sanford Rigidity Scale (4) were significantly slower than those scoring low in overcoming each of the three sets. On the other hand, subjects scoring high on the Dogmatism scale were significantly slower than those scoring low in integrating the three new sets into the problem solution after the three older sets had been overcome.

¹ This study was carried out while the senior author was a Faculty Research Fellow of the Social Science Research Council, 1951-1954.

The purpose of this study is not to demonstrate that dogmatism is a factorially pure measure independent of other measures or factors. Rather, its purpose is to determine empirically whether dogmatism as conceived and measured can be discriminated factorially, according to the considerations presented above, from the variables measured by the F scale (1), the E scale (1), and the Rigidity scale (4).

A secondary purpose of the present study has to do with the composition of the Dogmatism scale itself. The form employed in this study, as shown in Table 1, is composed of 43 items. Several of these items attempted to get at certain functional aspects of dogmatism. Thus, five items (Items 33-37) have to do with attitudes of acceptance or rejection of the self, and eight others (Items 36-43) have to do with a paranoid outlook on life. The question arises, are we justified in assuming that the self-rejection items, the paranoia items, and the other items more directly related to dogmatism belong together in the same scale? To investigate this question these three sets of items were scored separately and treated as independent variables in the analysis.

Finally, it was hypothesized that ideological dogmatism has its major motivational basis in anxiety. To the extent that this hypothesis is valid the three parts of the Dogmatism scale together with an Anxiety scale should emerge to define one of the factors in the analysis.

METHOD

A battery of ten scales was administered to 207 subjects (Ss) attending college in the metropolitan New York area. The scales were administered in group sessions under anonymous conditions. The Ss were students in elementary psychology courses. The items from the various scales were placed in random order in a single instrument and the Ss responded to each item on a 7-point agree-disagree scale from which the middle point had been excluded in order to avoid the piling up of responses at neutrality. For a more detailed description of the composition of the present sample and the over-all research context within which the present study is placed, see Rokeach (8).

The following scales were administered:

1. *Anxiety Scale*. Composed of 30 MMPI items found by Welch (9) to measure anxiety (e.g., "I have nightmares every few nights").

TABLE 1
THE DOGMATISM SCALE

1. A man who does not believe in some great cause has not really lived.
2. My blood boils whenever a person stubbornly refuses to admit he's wrong.
3. The worst crime a person could commit is to attack publicly the people who believe in the same thing he does.
4. It is when a person devotes himself to an ideal or cause that he becomes important.
5. Man on his own is a helpless and miserable creature.
6. In times like these, a person must be pretty selfish if he considers primarily his own personal happiness.
7. In a discussion people have a way of getting back at you by accusing you of quibbling and hair-splitting.
8. Communism and Catholicism have nothing in common.
9. There is nothing new under the sun.
10. To compromise with our political opponents is dangerous because it usually leads to the betrayal of our own side.
11. While I don't like to admit this even to myself, I sometimes have the ambition to become a great man, like Einstein, or Beethoven, or Shakespeare.
12. Unfortunately, a good many people with whom I have discussed important social and moral problems don't really understand what's going on.
13. It is sometimes necessary to resort to force to advance an ideal one strongly believes in.
14. It is better to be a dead lion than to be a live dog.
15. When it comes to differences of opinion in religion we must be careful not to compromise with those who believe differently from the way we do.
16. If a man is to accomplish his mission in life it is sometimes necessary to gamble "all or nothing at all".
17. A person who thinks primarily of his own happiness is beneath contempt.
18. It is better to be a dead hero than a live coward.
19. Nothing is impossible.
20. If I had to choose between happiness and greatness, I'd choose greatness.
21. I am not likely to speak to people until they speak to me.
22. To compromise with our political opponents is to be guilty of appeasement.
23. It's all too true that most people just won't practice what they preach.
24. Once I get wound up in a heated discussion I just can't stop.
25. The United States and Russia have just about nothing in common.
26. In a heated discussion I generally become so absorbed in what I am going to say that I forget to listen to what the others are saying.
27. To one who really takes the trouble to understand the world he lives in, it's a relatively easy matter to predict future events.
28. To achieve the happiness of mankind in the future it is sometimes necessary to put up with injustices in the present.
29. The present is all too often full of unhappiness. It is the future that counts.
30. It is by returning to our forgotten and glorious past that real social progress can be achieved.
31. I am afraid of people who want to find out what I'm really like, for fear they'd be disappointed in me.
32. I feel self-conscious when I'm with people who have a superior position to mine in school or business.
33. I'd like it if I could find someone who would tell me how to solve my personal problems.
34. My hardest battles are with myself.
35. At times I think I am no good at all.
36. All too many people are failures and it is the system which is responsible for this.
37. It is only natural for a person to have a guilty conscience.
38. It is generally safer to trust nobody.
39. Much of the time I feel as if I have done something wrong or evil.
40. I have often felt that strangers were looking at me critically.
41. If given the chance I would do some things that would be of great benefit to the world.
42. I have often felt that people say insulting and vulgar things about me.
43. I am sure I am being talked about.

2. *Paranoia Subscale*. Composed of Items 36-43 in Table 1 (e.g., "It is safer to trust nobody").²

3. *Self-Rejection Subscale*. Composed of Items 31-35 in Table 1 ("I am afraid of people who want to find out what I'm really like, for fear they'd be disappointed in me").³

4. *Dogmatism Subscale*. Composed of the first 30 items in Table 1 (e.g., "A man who does not believe in some great cause has not really lived").⁴

5. *Authoritarianism (F Scale)*. Composed of 29 items (e.g., "Obedience and respect for authority are the most important virtues children should learn") (1, Form 45-40, pp. 255-257).

6. *Rigidity Scale*. Composed of 22 items (e.g., "I wish people would be more definite about things") (4).⁵

of increasing its reliability. A slightly modified version of this scale was employed in a study by Rokeach, McGovney, and Denny (7). The final form of the Dogmatism scale is composed of 40 items and is reported fully in a separate publication (8).

⁵ The writers are grateful to Drs. Gough and Sanford for their courtesy in permitting use of this scale. This scale is also included in Gough's California Psychological Inventory (copyright 1951). Some sample

² Items 38-42 are taken from the MMPI (4).

³ Items 33, 34, 35 are taken from Berger (2), Items 36 and 37 from MMPI (4).

⁴ The corrected reliability of this scale is .75. Since this study was completed the Dogmatism scale has undergone further revision, primarily for the purpose

TABLE 2
CORRELATION MATRIX
($N = 207$)

	1	2	3	4	5	Variable					10	r^*_{tt}
						6	7	8	9			
1. Anxiety		.65	.71	.33	.17	.14	.13	-.13	.22	-.05		.90
2. Paranoia			.52	.41	.30	.29	.20	-.08	.31	-.03		.60
3. Self-rejection				.30	.14	.07	.10	-.14	.11	-.02		.64
4. Dogmatism					.64	.62	.52	.23	.12	.33		.75
5. Authoritarianism (F Scale)						.69	.63	.43	-.08	.56		.84
6. Rigidity							.54	.35	.02	.40		.83
7. Ethnocentrism (E Scale)								.44	-.22	.62		.72
8. Conservatism (P.E.C.)									-.39	.60		.62
9. Left Opinionation											-.51	.74
10. Right Opinionation												.77

* Odd-even reliability estimates of the scales corrected for length.

TABLE 3
FACTOR LOADINGS

Variable	Rotated Loadings			Centroid Loadings			R^2
	I	II	III	I	II	III	
1. Anxiety	.77	-.25	.27	.442	-.708	.173	.727
2. Paranoia	.72	-.14	.26	.468	-.600	.134	.597
3. Self-rejection	.69	-.29	.37	.412	-.656	.313	.698
4. Dogmatism	.46	.21	.62	.708	-.139	.341	.637
5. Authoritarianism (F scale)	.27	.48	.66	.784	.185	.297	.737
6. Rigidity	.23	.32	.71	.680	.123	.421	.652
7. Ethnocentrism (E scale)	.21	.59	.47	.728	.273	.099	.614
8. Conservatism (P.E.C.)	-.07	.69	.19	.487	.519	-.129	.523
9. Left opinionation	.17	-.63	.26	-.169	-.500	.468	.498
10. Right opinionation	.03	.85	.19	.642	.549	-.227	.765

7. *Ethnocentrism (E Scale)*. Composed of 10 items about Negroes and other minority groups, but excluding items about Jews. (e.g., "Manual labor and unskilled jobs seem to fit the negro mentality and ability better than more skilled or responsible work") (1, Item 32, p. 128; Items B1-B4, and C1-C5, p. 142).

8. *Political-Economic Conservatism (PEC Scale)*. Composed of five items (e.g., "In general, full economic security is bad; most men wouldn't work if they didn't need the money for eating and living.") (1, Form 45-40, p. 169).⁶

9. *Left Opinionation Scale*. Composed of 20 items designed to measure degree of intolerance in a left

items are: "I wish people would be more definite about things;" "Once I have made up my mind I seldom change it;" "I always see to it that my work is carefully planned and organized;" "I like to have a place for everything and everything in its place."

⁶ However, the P.E.C. item "Most government controls over business should be continued even though the war is over" was revised to read: "Most government controls over business should be continued and even strengthened."

direction (e.g., "It's just plain stupid to keep Red China out of the United Nations").⁷

10. *Right Opinionation Scale*. Composed of 20 items designed to measure degree of intolerance in a right direction (e.g., "It's the agitators and left wingers who are trying to get Red China into the United Nations").

RESULTS AND DISCUSSION

The scores derived from the several scales were intercorrelated by the Pearson product-moment method and the resulting coefficients as well as the corrected odd-even reliability estimates are presented in Table 2. The correlation matrix was factored by the centroid

⁷ A fuller discussion of the total Opinionation scale and the several subscales thereof is presented elsewhere (8). Our purpose for including the Left and Right Opinionation subscales in this study was to find out whether these measures, along with the P.E.C. scale, would emerge together on a liberalism-conservatism factor.

method and the three factors extracted were rotated to approximate orthogonal simple structure (3). The rotated loadings, centroid loadings, and communalities for the ten variables are shown in Table 3.

The variables with the highest loadings on the first rotated factor are the Anxiety, Paranoia, and Self-Rejection. Dogmatism also has a substantial loading on this factor. Left and Right Opinionation and Political-Economic Conservatism have the highest loadings on the second rotated factor. The E scale and the F scale also have substantial loadings toward the "right" end of the bipolar factor. With respect to the third rotated factor it is seen that Rigidity, the F scale, and Dogmatism have the highest loadings and the E scale has a moderate loading on it. While it is difficult to attach definite labels to these three factors, Factor I may be tentatively identified as representing anxiety, Factor II liberalism-conservatism, and Factor III rigidity-authoritarianism.

More important for the purposes of this study than the clear-cut identification of these factors is any assistance they may be in clarifying the issues stated at the beginning of the paper.

1. *Is the Dogmatism subscale discriminable from the F scale, the Rigidity scale and the E scale?* It is seen that while both dogmatism and authoritarianism have their major loadings on Factor III, the pattern of loadings for these two variables differs on the first two factors. The F scale has a substantial loading toward the right end of the liberalism-conservatism factor (Factor II) and a negligible loading on the anxiety factor (Factor I) in contrast to the Dogmatism subscale which has the opposite profile on these two factors. These findings suggest that even though Dogmatism and F are correlated with each other ($r = .64$) Dogmatism is relatively independent of the left-right continuum while the F scale is not. The situation for the two scales is the reverse with respect to the anxiety factor, with dogmatism showing a substantially higher loading than authoritarianism on that factor.

The results bearing on the Rigidity scale indicate that it has its major loading on Factor III. It has minor loadings on the first two factors and comes closest to being a pure representative of the third factor. As for the E

scale, it has its largest loading on Factor II (liberalism-conservatism) and also a moderately high loading of Factor III (rigidity-authoritarianism).

It is thus seen that Dogmatism differs from the E and F scales in that it is independent of the left-right continuum and has a strong component on the anxiety factor, whereas the other two scales are oriented toward the right end of the liberalism-conservatism factor and have only negligible loadings on the anxiety factor. Dogmatism can be distinguished factorially from rigidity in that the latter has its only substantial loading on the rigidity-authoritarianism factor.

2. *Are dogmatism, self-rejection, and paranoia factorially similar?* It is seen in Table 3 that these three variables do indeed have substantial loadings, along with anxiety, on Factor I. Thus, there is good supporting evidence that the eight items having to do with paranoid and the five items having to do with self-rejecting tendencies do contribute to the anxiety aspect of the total Dogmatism scale. At the same time it is noted that there is a difference between the Dogmatism subscale (30 items) on the one hand and the paranoid and self-rejection subscales on the other. The Dogmatism subscale has a relatively high loading on Factor III (rigidity-authoritarianism), while the paranoia and self-rejection variables do not.

3. *What is the relation between dogmatism and anxiety?* It is seen further that anxiety has a very heavy loading on Factor I and relatively minor loadings on Factors II and III. In accord with our hypothesis anxiety is seen to emerge on a single factor along with the three parts of the total Dogmatism scale already referred to.

SUMMARY AND CONCLUSIONS

A factor analysis was conducted using a sample of 207 New York college students, of scales representing the following ten variables: anxiety, paranoia, self-rejection, dogmatism, authoritarianism, rigidity, ethnocentrism, liberalism-conservatism, left opinionation, and right opinionation. The following three major hypotheses were investigated: (a) that dogmatism as conceived and measured is factorially discriminable from authoritarianism, ethnocentrism, and rigidity; (b) that dogmatism,

paranoia and self-rejection are factorially similar; (c) that dogmatism, paranoia, and self-rejection would emerge, together with anxiety, on a single factor.

On the whole it can be concluded from the data that the three hypotheses are confirmed. Factor I is composed primarily of anxiety, paranoia, and self-rejection, with dogmatism also having a substantial loading on this factor. Factor II is primarily a liberalism-conservatism factor being composed primarily of left and right opinionation, the Political-Economic Conservatism Scale, and to a lesser extent, ethnocentrism and the F scale. Factor III includes primarily rigidity, the F scale, and dogmatism.

From all the preceding we conclude that dogmatism, as conceived and represented, measures something similar to authoritarianism but independently of the left-right dimension (Factor II). It is also discriminable from rigidity, and ethnocentrism. Finally, it seems to be related to anxiety.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778-782.
3. FRUCHTER, B. *Introduction to factor analysis*. New York: Van Nostrand, 1954.
4. GOUGH, H. G., & SANFORD, R. N. Rigidity as a psychological variable. Berkeley: Univer. of Calif. Inst. of Pers. Assessment and Res., 1952 (mimeographed).
5. HATHAWAY, S. R., & MCKINLEY, J. C. *The Minnesota Multiphasic Personality Inventory*. New York: Psychological Corp., 1943.
6. ROKEACH, M. The nature and meaning of dogmatism. *Psychol. Rev.*, 1954, **61**, 194-204.
7. ROKEACH, M., MCGOVNEY, W. C., & DENNY, M. R. A distinction between dogmatic and rigid thinking. *J. abnorm. soc. Psychol.*, 1955, **51**, 87-93.
8. ROKEACH, M. Political and religious dogmatism: An alternative to the authoritarian personality. *Psychol. Monogr.*, **70**, No. 18 (Whole No. 425).
9. WELCH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16**, 65-72.

Received January 18, 1956.

PERFORMANCE AS A FUNCTION OF MOTIVE STRENGTH AND EXPECTANCY OF GOAL-ATTAINMENT

JOHN W. ATKINSON AND WALTER R. REITMAN¹

University of Michigan

THE results of recent studies that have related thematic apperceptive measures of achievement (7) and affiliation motives (3) to performance have been consistent with a principle of performance adapted from Tolman's (11) expectancy theory: the goal-directed action tendency is a joint function of the strength of the motive and of the expectancy aroused by situation cues that performance is instrumental to attainment of the goal of the motive (2). Our purpose is to elaborate this idea as it applies to the relation of performance to measures of individual differences in human motives and to present an experimental test of a prediction when more than one motive and one expectancy are involved.

Earlier studies (1, 4, 5, 6, 12) have shown a positive relationship between performance and n-Achievement scores obtained from thematic apperceptive stories when the cues of the performance situation arouse the expectancy that a feeling of personal accomplishment will accompany a good performance. On the other hand, when the cues for performance are deliberately manipulated so that Ss are given no reason to expect they will experience pride in accomplishment, there is no relationship between performance and n-Achievement scores (1, 4, 5). On two occasions (4, 5), the E's explicit appeal for cooperation in situations not related to achievement has apparently served to engage another motive. Performance level, while unrelated to strength of n Achievement, has been found positively related to n Affiliation. French (5)² has also shown that when an incentive unrelated to either n Achievement or n Affiliation is offered, no

systematic relationship is found between performance and either of these motives.

We conceive a *motive* as a latent disposition to strive for a particular goal-state or aim, e.g., achievement, affiliation, power. The strength of a particular motive is assessed by thematic apperception under *neutral*³ conditions. The term *motivation* can then be used to designate the aroused state of the person that exists when a motive has been engaged by the appropriate expectancy, i.e., an expectancy that performance of some act is instrumental to attainment of the goal of that motive. If more than one of an individual's motives are engaged by expectancies that the same act will lead to several different goals, the *total motivation* for performance of that act will be the sum of the contributions made by the particular motives which have been engaged. An overdetermined action tendency, i.e., one which will serve to satisfy more than one motive, is likely to be relatively strong.

This conception implies that a positive relationship between a particular motive (e.g., n Achievement) and performance can be maximized by engaging that motive and no others in performance and can be minimized either (a) by failing to engage that motive at all as in (1, 4, 5) or (b) by systematically engaging other motives in performance *as well*, as suggested in (5, 12). In the latter case, the person who is weak in the motive the E may have measured may be strong in other unmeasured (ulterior) motives that are engaged in performance of the same act.

The missing link among the studies already accomplished is a demonstration that the relationship between n Achievement and performance can be substantially reduced even *when the achievement expectancy is explicitly aroused* by systematic arousal of other motives to perform the same act. This should occur, assuming an asymptote for performance, when

¹ This study was accomplished during the term of a Social Science Research Council fellowship to JWA and a National Science Foundation fellowship to WRR. We should like to acknowledge, also, assistance provided by the Center for Advanced Study in The Behavioral Sciences.

² French's motive scores were obtained through content analysis of responses made to a series of single-sentence descriptions of behavior which Ss were asked to explain as a "Test of Insight." The categories of response are directly comparable to those employed in the analysis of TAT stories.

³ A *neutral* condition is one in which no experimental attempt is made either to arouse the motive or to create an especially relaxed state prior to the writing of stories (see 7).

the total motivation of a low *n*-Achievement group approximates that of a high *n*-Achievement group. Our experimental task, then, is to engage the achievement motive in two different situations by instructions that are known to arouse it. But then we must manipulate other features of the two situations so that expectancies of attaining other goals are relatively weak or absent in one situation and strong in the other.

Taking a lead from Wendt (12), we designed a group test situation in which two examiners would stalk about the room with watches in hand, pacing performance with a signal every minute and, to the best of their ability, casting disapproving glances at anyone who seemed to be gazing around or not taking the task seriously. These innovations, in addition to the usual type of achievement-orienting instructions, were introduced to engage the affiliation motive, the desire to please the experimenter. Finally, in an appeal to the motivational structure of persons who might be low in both *n* Achievement and *n* Affiliation, a substantial monetary reward (\$5.00) was offered to the person getting the best score on the test.

In another condition, the same achievement-orienting instruction was used to engage the achievement motive, but to minimize the arousal of other goal-expectancies, the *Ss* were assigned individually to small test rooms and left to work alone without even the competitive cues provided by the sight and sound of other persons working. In this condition, the *E* was not present during performance, and there was no monetary incentive.

The performance level of persons who are high in *n* Achievement should be substantially higher than that of persons who are low in *n* Achievement in the group given achievement-orienting instructions and then left on their own to work without expectancies of satisfying other motives. The difference should be substantially reduced (or removed completely) in the other group given achievement-orienting instructions but in whom expectancies of attaining other goals (affiliation and money) were also aroused. No relationship is expected between *n* Affiliation and performance in either condition. In one group, only the expectancy of achievement was aroused, while in the other, expectancies of achievement and a monetary reward were also present to con-

found the simple relationship to *n* Affiliation that might otherwise have resulted from the introduction of affiliation-related cues.

METHOD

Subjects were 93 male college students selected from a group of several hundred introductory psychology students for whom measures of both achievement and affiliation motives were available.

Measurement of achievement and affiliation motives. Eight four-minute imaginative (TAT-type) stories were written under neutral (7) classroom conditions one to three months before the experimental sessions. The results of a methodological study run concurrently indicate that the first four stories in a series of eight, irrespective of pictures used, provide the greatest predicative validity for *n* Achievement (9). The motive scores of the present study were therefore based only on the first four stories. The rank-order score-rescore reliability was .88 for *n* Achievement (7) and .89 for *n* Affiliation (3). The distributions of scores of the larger population from which the sample was drawn were divided at the median to yield high and low groups with respect to *n* Achievement and *n* Affiliation.

Experimental conditions. Forty-nine *Ss* in groups of six to ten were assigned to the *Achievement-Orientation* condition. The *E* emphasized the importance of doing well on the tasks by alluding to them as similar to tests of important abilities (executive capacity, leadership, intellectual alertness) developed by psychologists in the past. The *Ss* were urged to do their best and were told, "your work will be taken as the full measure of your ability" (see 1). They were also told that they would be assigned to individual test rooms to prevent distractions.

Following this general achievement-orienting instruction, and before being assigned to individual test rooms, *Ss* were given a short story to read and told to memorize its content for later recall. This task, part of another study not reported here,⁴ took three minutes. It was introduced to take advantage of the aroused motivational state of the *Ss* on the assumption that it would in no way affect the performance of *Ss* on the subsequent tasks of the present experiment. As it turned out, this assumption was probably incorrect.

The *Ss* were then assigned to individual test rooms. The first task, which consisted of drawing *Xs* inside of circles, was explained to them. When the test was delivered to each *S*, he was told to begin and urged, "do your best," or, "see how good you are at this one." The *E* then noted on a stop watch the exact time the *S* started working and recorded this time in seconds immediately after leaving the room. He then allowed 10-15 seconds to elapse before delivering a test to the next *S* in an adjacent room to make it possible for him to stop each *S* exactly 14 minutes after his own starting time. The *E* entered each room and stopped each *S* within five seconds of 14 minutes after he had started.

Instructions for the second task, a set of simple

⁴ Reitman, W. R., Motivation and recall of meaningful material under achievement-orientation. (Unpublished paper) University of Michigan, 1955.

three-step arithmetic problems, were then distributed and explained. The delivery of the second task followed the same procedure. Each *S* was told, "this one will demand a little more brain power," or, "this one will be a lot harder," when the test itself was delivered. The same procedure of timing *Ss* was employed.

Forty-four *Ss* were assigned to the *Multi-Incentive* condition. The *Ss* in this condition met in groups of 10-20 and stayed together in a large classroom for performance of both tasks. The same general achievement-orienting instruction was given followed by the three-minute memorization task. The *E* then said, "Since we are interested in seeing your very best performance when you are actually putting out, we are going to award a prize of \$5.00 to the person having the highest score on each of the tests. So you could stand to walk off with \$10.00 for your efforts today." Finally, *Ss* were told they would be asked to skip an item from time to time throughout the test period.

Throughout the performance of each test, the *Es* walked about the room, stop watches in hand, as if evaluating the seriousness of each *S's* effort. At the end of each minute, one of the *Es* called out, "skip," at which point the *Ss* stopped, skipped an item, and moved on to the next one.

Fourteen minutes were given for each task, and the monetary incentive with a reminder, "do your best," was repeated before the second task which was also paced.

The tasks. The first task consisted of a 15-page booklet containing row after row of $\frac{1}{2}$ inch circles in which the *S* was instructed to draw *Xs* (4). The second task, arithmetic problems, required solution of a series of three-step problems. There were 14 pages, each containing 24 problems. The operations consisted of adding or subtracting three digits, remembering the solution while adding or subtracting another set of three digits, and then finally adding or subtracting the two solutions (12).

RESULTS

Drawing Xs in circles. The mean number of *Xs* drawn in circles was 1255.5 for the *Multi-Incentive* condition ($N = 44$) and 1149.1 for the *Achievement-Oriented* condition ($N = 51$). The difference of 106.4 between the two conditions is significant ($t = 3.62$, $p < .01$), but there was no difference between high and low *n-Achievement* groups within either condition.

We suspect that performance of the first task was confounded by continued attempts to practice the memory material that had been introduced at the beginning of the test period. When queried about possible effects of the memory task after the experimental session, a number of *Ss* indicated that they had continued to practice recall during the first task. In light of the results of one earlier study (4) and another in preparation that show the *Xs-in-circles* task to be a sensitive measure of

effort and significantly related to differences in *n-Achievement* under appropriate conditions, we assume that many of the *Ss* felt it more important to concentrate on the memory material than to do their best at the relatively simple motor task that followed. Continued rehearsal of the memory material would be sufficiently distracting to influence the level of performance.

The achievement motive and arithmetic performance. The results for the second task are presented in Table 1. The arithmetic problems, which demanded continued attention and concentration, would not encourage continued attempts at memorization. We therefore view the results presented in Table 1 as a less confounded test of the initial hypothesis.

In the *Achievement-Oriented* condition, the performance level of the high *n-Achievement* group was significantly higher ($p < .01$) than that of the low *n-Achievement* group on both number of solutions attempted and number correct. In the *Multi-Incentive* condition, there was no difference between the performance levels of the two motivation groups. The difference between the differences is 19.8 for attempted solutions and 15.9 for correct solutions. Both are significant in the direction predicted ($ts = 2.30$ and 1.84 respectively, $p < .05$). In other words, the relationship between achievement motive and arithmetic performance was eliminated by systematically

TABLE 1
ARITHMETIC PERFORMANCE AS A FUNCTION OF
ACHIEVEMENT MOTIVE AND EXPERIMENTAL
CONDITION

Achievement Motive	Condition					
	Achievement Orientation			Multi-Incentive		
	<i>N</i>	Attempted Solutions	Correct Solutions	<i>N</i>	Attempted Solutions	Correct Solutions
High	21	M 78.1 σ 24.8	71.6 24.1	24	M 67.1 σ 19.6	60.3 19.3
Low	30	M 60.3 σ 15.7	55.5 16.4		M 69.1 σ 22.3	60.1 23.0
Diff (H-L)		17.8	16.1		-2.0	.2
σ diff.		5.93*	5.96		6.23	6.26
<i>t</i>		3.00	2.70		n.s.	n.s.
<i>p</i>		.01	.01			

* σ diffs. derived from estimate of within group variance, $df = 92$.

engaging other motives in the same performance.

An unanticipated fault in our method of pacing is chiefly responsible for the failure of over-all performance level to increase in the Multi-Incentive condition. When *E* called out, "skip," at the end of each minute, *Ss* had to stop working on partially solved problems and move immediately to the next one. On the earlier task, they had merely to skip one circle. Since there were 13 such interruptions, the performance of the high *n*-Achievement group would suffer a decrement if it were already nearly asymptotic in the Achievement-Oriented condition. And the net gain of the low *n*-Achievement group, which should profit most by the arousal of other motives, would also be depressed. The high *n*-Achievement group actually dropped 11.0 in number of attempts and 11.3 in number correct. The low *n*-Achievement groups showed net gains of 8.8 in number of attempts and 4.6 in number correct. The extent to which the decrease of the high *n*-Achievement group may be a real decrement accompanying very intense motivation is, unfortunately, indeterminate.

Joint effect of motivation and aptitude on arithmetic performance. Lowell (6, 7) found no relationship between *n* Achievement and quantitative aptitude among college students. His results can be checked in the present study.

Quantitative Aptitude Scores (*Q* scores) from the American Council of Education Test were available for 38 *Ss* in the present Achievement-Oriented condition. The data for 20

additional *Ss* for whom *Q* scores could be obtained were added. These 20 *Ss* had been run under the same Achievement-Oriented condition but as part of a methodological study (9). They were divided into high and low *n*-Achievement groups according to the median score of a much larger group who had taken a different and somewhat less sensitive form of the thematic apperceptive test of motivation. The obtained distribution of 58 *Q* scores (percentile rank using local norms) was divided into thirds to provide High, Middle, and Low aptitude groups.

The mean *Q* score was 64.6 for this high *n*-Achievement group ($N = 29$) and 57.2 for the low *n*-Achievement group ($N = 29$). The difference, 7.4, is insignificant ($t = 1.05$), confirming Lowell's finding.

The mean number of arithmetic problems attempted by this high *n*-Achievement group ($N = 29$), irrespective of aptitude, was 71.2; the mean for the low *n*-Achievement group ($N = 29$) was 61.7. The difference, 9.5, while less than the comparable difference in the experiment proper (Table 1) as a consequence of the less sensitive measure of achievement motive for the 20 additional *Ss*, is nevertheless significant ($t = 2.06$, $p < .05$). Similarly, the performance levels of High ($N = 18$), Middle ($N = 22$), and Low ($N = 18$) aptitude groups, independent of motive strength, are significantly different as normally expected ($F = 3.23$, $df = 2$ and 55 , $p < .05$).

Table 2 shows that the highly motivated groups were uniformly higher in performance

TABLE 2

ARITHMETIC PERFORMANCE AS A JOINT FUNCTION OF ACHIEVEMENT MOTIVE AND QUANTITATIVE APTITUDE (Achievement-Oriented Condition)

Quantitative Aptitude (Percentile Rank)	Achievement Motive						Diff.	σ diff.	<i>t</i>	<i>p</i>
	High			Low						
	<i>N</i>	Attempted Solutions	Correct Solutions	<i>N</i>	Attempted Solutions	Correct Solutions				
High 80-99th	11	76.9	70.5	7	72.4	68.5	4.5	8.4	n.s.	
Middle 53-75th	10	65.7	61.2	12	59.8	54.7	2.0	8.5	n.s.	
Low 3-48th	8	70.4	62.8	10	56.4	49.0	5.9	7.4	n.s.	
							6.5	7.5	n.s.	
							14.0	8.2	1.71	.05**
							13.8	8.3	1.66	.05**
Diff. (H-L)		6.5	7.7		16.0	19.6				
σ diff.		8.1*	8.2		8.6	8.7				
<i>t</i>		—	—		1.86	2.25				
<i>p</i>		—	—		.05**	.05**				

* σ diffs. derived from estimates of within group error, $df = 52$.

** One-tailed tests.

than the less motivated groups at the three levels of aptitude. But the effect of motivation was most pronounced in the group that was lowest in aptitude. Here the difference between high and low *n*-Achievement groups in number of attempted solutions was 14.0 ($t = 1.71$, $p < .05$ in predicted direction). It is less than half as large at the middle level of aptitude (diff. = 5.9, σ diff. = 7.4) and at the high level of aptitude (diff. = 4.5, σ diff. = 8.4). The difference between the differences of High and Low aptitude groups, 9.5, is not significant (σ diff. = 11.8). The result nevertheless suggests that persons who are low in aptitude profit most by strong motivation. The trends for number of correct solutions are very similar, as Table 2 shows, but the differences are somewhat smaller.

Table 2 shows a comparable result regarding the effect on performance of differences in aptitude, viz., the effect is greatest when motivation is low. The difference between High and Low aptitude groups also low in *n* Achievement was 16.0 for attempted solutions (σ diff. = 8.6, $t = 1.86$, $p < .05$ in expected direction). The comparable difference in the group who were highly motivated to achieve is only 6.5 (σ diff. = 8.1). The result is essentially the same, though in this case slightly more pronounced, for number of correct solutions. In neither case is the difference between the differences significant.

The affiliation motive and arithmetic performance. No systematic relationship between *n* Affiliation and performance is expected in either of the experimental conditions. Table 3 shows that none of the differences is significant when a two-tailed t -test is made. The tendency for the low *n*-Affiliation group to perform at a somewhat higher level under Achievement-Oriented can probably be attributed to the low negative correlation ($-.20$ to $-.25$) between *n* Achievement and *n* Affiliation when scores are taken from the same four stories (9). In light of this, the *positive* trend between *n* Affiliation and performance in the Multi-Incentive condition, particularly for correct solutions ($p = .10$), may mean that we were not completely successful in our attempt to eliminate the relationship between performance and a particular motive by systematically arousing other motives. The positive trend, while statistically very tenuous, suggests that the expectancy of pleasing the *E*s by working

TABLE 3
ARITHMETIC PERFORMANCE AS A FUNCTION OF
AFFILIATION MOTIVE AND EXPERIMENTAL
CONDITION

Affiliation Motive	Condition					
	Achievement Orientation			Multi-Incentive		
	<i>N</i>	Attempted Solutions	Correct Solutions	<i>N</i>	Attempted Solutions	Correct Solutions
High	28	M 63.3 σ 15.8	57.7 15.9	17	M 75.1 σ 22.5	68.3 22.9
Low	23	M 73.0 σ 26.4	67.5 25.8	28	M 66.1 σ 18.6	57.4 18.3
Diff (H-L)		-9.7	-9.8		9.0	10.9
σ diff.		6.0*	5.9		6.5	6.3
t		1.62	1.66		1.38	1.72
p		n.s.	n.s.		n.s.	.10

* σ diffs. derived from estimate of within group error, $df = 92$.

hard was most salient in the Multi-Incentive condition. Comparison of changes in performance between the two conditions supports this view, as *S*s in the Multi-Incentive condition high in *n* Affiliation and low in *n* Achievement ($N = 8$) averaged 16.7 more attempted solutions than *S*s with comparable motive scores in the Achievement-Oriented condition ($N = 20$); all other *S*s in the Multi-Incentive condition ($N = 37$) averaged 5.8 fewer attempted solutions than the comparable Achievement-Oriented *S*s ($N = 31$). The difference, 22.5, is significant (σ diff. = 10.0, $t = 2.25$, $p < .05$). The correct solutions result is very similar. The greatest increase in performance due to the presence of the *E*s and the monetary reward was among *S*s highly motivated to affiliate but weak in motivation to achieve.

DISCUSSION

Genotype and phenotype. We interpret the results as consistent with the conception of motivation and behavior elaborated in the introductory section. The distinction between strength of particular motives (latent dispositions) and strength of motivation to perform a particular act (the sum of all aroused motives) highlights the genotype-phenotype distinction that is often made in treatments of motivation and behavior. An obvious implication of the results of this and earlier, related studies (1, 4, 5, 12) is that psychologists should stop expecting to find measures of *particular* motives and performance criteria like academic

performance, leadership, etc., to be more highly related than they have been found in the past (for example 8). From the present viewpoint, most performance criteria in real-life situations can be considered overdetermined, i.e., acts or accomplishments for which several of the individual's motives have been engaged.

The role of the situation. The present formulation calls for a good deal more thought than is commonly given the question of conceptualizing and ultimately measuring the properties of the situation in which behavior occurs. Our difficulty in attempting to create an experimental situation that would make the expectancies of achievement, affiliation, and monetary prize nearly equal in salience, as evidenced by the positive trend between *n* Affiliation and performance in the Multi-Incentive condition, illustrates the need for adequate methods of assessing situations in terms of the expectancies of goal-attainment they provoke. Rotter (10) and his colleagues have begun to make substantial strides in this direction.

SUMMARY

Particular motives, (Achievement, Affiliation, Power, etc.) are conceived as latent dispositions to strive for certain goal states (aims) that are engaged in performance of an act when the cues of the situation arouse the expectancy that performance of an act is instrumental to attainment of the goal of the motive. The total *motivation* to perform the act is conceived as a summation of strengths of all the *motives* that have been aroused by appropriate expectancies of goal-attainment cued-off by the situation. The relationship of achievement motive (as measured by imaginative TAT stories) to performance is shown to be significantly positive when the expectancy that performance is instrumental to producing a feeling of pride in accomplishment is aroused and few if any other expectancies of goal-attainment are aroused. When motives for other goals

(affiliation and money) are *also* aroused by deliberate manipulation of situational cues that activate the *Ss'* expectancies, there is no relationship between achievement motive and performance. The joint effect of quantitative aptitude and achievement motive on arithmetic performance is illustrated in a subsidiary analysis of the data.

REFERENCES

1. ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381-390.
2. ATKINSON, J. W. Explorations using imaginative thought to assess the strength of human motives. In *Nebraska symposium on motivation*, 1954. Lincoln: Univer. of Nebraska Press, 1954.
3. ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.
4. ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349-363.
5. FRENCH, E. G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232-236.
6. LOWELL, E. L. The effect of need for achievement on learning and performance. *J. Psychol.*, 1952, **33**, 31-40.
7. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
8. MORGAN, H. A. Measuring achievement motivation with "Picture Interpretations." *J. consult. Psychol.*, 1953, **17**, 289-292.
9. REITMAN, W. R., & ATKINSON, J. W. Methodological problems in the use of thematic apperceptive measures of human motivation. In J. W. Atkinson (Ed.), *The assessment of human motives*. Princeton: Van Nostrand, in press.
10. ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation*, 1955. Lincoln: Univer. of Nebraska Press, 1955.
11. TOLMAN, E. C. Principles of performance. *Psychol. Rev.*, 1955, **62**, 315-326.
12. WENDT, H. W. Motivation, effort, and performance. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955.

Received January 20, 1956.

COGNITIVE STRUCTURE AND ATTITUDINAL AFFECT¹

MILTON J. ROSENBERG

Yale University

UNDERSTANDING of the related processes of attitude learning and attitude change will probably be advanced by the investigation of structural relationships between attitudes and beliefs about the objects of attitudes. The present research is an attempt to verify a set of hypotheses about such relationships. These hypotheses have much in common with some that have already been presented by Cartwright (2), Hilliard (3), Smith (8), Tolman (9), and Woodruff (11, 12). However, they differ from these earlier formulations in attempting to delineate more explicitly certain variables that are assumed to covary with attitude (here defined as *relatively stable affective response to an object*). One of these variables is the intensity of a person's values. A second is the perceived importance of the attitude object in leading to or blocking the attainment of his values.

The general theoretical view underlying the present study includes the following points: (a) When a person has a relatively stable tendency to respond to a given object with either positive or negative affect, such a tendency is accompanied by a *cognitive structure* made up of beliefs about the potentialities of that object for attaining or blocking the realization of valued states; (b) the sign (positive or negative) and extremity of the affect felt toward the object are correlated with the content of its associated cognitive structure. Thus strong and stable positive affect toward a given object should be associated with beliefs to the effect that the attitude object tends to facilitate the attainment of a number of important values, while strong negative affect should be associated with beliefs to the effect that the attitude object tends to block the attainment of important values. Similarly, moderate positive or negative affects should be associated with beliefs that

relate the attitude object to less important values or, if to important values, then with less confidence as to the existence of clear-cut instrumental relationships between the attitude object and the values in question.

From this view, three specific hypotheses were formulated for experimental test:

1. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of the products obtained by multiplying the rated importance of each value associated with that object by the rated potency of the object for achieving or blocking the realization of that value.

2. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of his ratings of the potency of that object for achieving or blocking the realization of his values (when the importance of these values is held constant).

3. The degree and sign of affect aroused in an individual by an object (as reflected by the position he chooses on an attitude scale) vary as a function of the algebraic sum of his ratings of the importance of the values whose attainment or blocking he perceives to be affected through the instrumental potency of that object (when the instrumental potency of that object for attaining or blocking each of these values is held constant).

METHOD

Subjects and Attitude Measure

One hundred and twenty Ss, recruited from undergraduate courses at the University of Michigan and the Ypsilanti State Teachers College, took an attitude questionnaire that contained, among other items, one dealing with the issue of "whether members of the Communist Party should be allowed to address the public." Each S checked his first choice among five alternative statements. Seventeen Ss chose the alternative indicating extreme opposition to allowing members of the Communist Party to address the public, 31 Ss the alternative indicating moderate opposition, 44 Ss that indicating moderate approval, and 25 Ss that indicating extreme approval. Only three Ss chose the

¹ This paper is based upon part of a doctoral dissertation submitted to the University of Michigan. The study was supported by the U. S. Air Force, Human Resources Research Institute, under contract #AF33-(038)-26646. The advice of Professor Helen Peak was of great value during all phases of the investigation.

alternative indicating "neutrality" on the issue, and, because of their small number, were excluded from the research population. In a retest of 95 Ss on this measure after an interval of at least two months, a reliability coefficient of .72 was obtained ($p < .001$).

Value Measures

Three to five weeks after the administration of the attitude measure two card-sorting tasks were administered individually. These tasks required S to categorize each of a group of value items in terms of (a) *value importance*, i.e., its importance to him as a "source of satisfaction" and (b) *perceived instrumentality*, i.e., his estimate as to whether, and to what extent, the

TABLE 1
VALUE ITEMS USED IN CARD-SORTING TASKS*

1. People sticking to their own groups.
2. People looking out for the welfare of others.
3. Being looked up to by others.
4. Change and variety; having new kinds of experience.
5. Sticking to a difficult task; solving difficult problems.
6. Making one's own decisions.
7. People being strongly patriotic.
8. Serving the interests of the group to which one belongs.
9. Giving expression to feelings of anger or hostility.
10. Keeping promises made to others.
11. Having one's family approve of one's views.
12. The uncompromising administration of punishment to anyone who deserves it.
13. Having interesting work to do.
14. Having power and authority over people.
15. Being well-informed about current affairs.
16. Having the value of property well protected.
17. People of different backgrounds getting to know each other better.
18. All human beings having equal rights.
19. Being good-looking; having attractive face, body, or clothes.
20. Having a steady income.
21. Believing in a relationship between the individual and some higher spiritual power.
22. America having high prestige in other countries.
23. Being liked or loved by the opposite sex or associating with the opposite sex.
24. Getting advice on important problems.
25. People having the right to participate in making decisions which will affect them.
26. Achieving superiority over others in such things as knowledge, work, or sports.
27. Complying with the wishes of persons in authority.
28. Being like others in general; having the same interests.
29. People being well educated.
30. People having strict moral standards.
31. The open expression of disagreement between people.
32. Everyone being assured of a good standard of living.
33. Being allowed to maintain the privacy of one's opinions and beliefs.
34. Being with other people; socializing.
35. Letting others make their own decisions.

* This list does not include the "salient" value items specially constructed for each S.

value in question would tend to be achieved or blocked through the "policy of allowing members of the Communist Party to address the public."

The pack of value-cards included 35 items constructed in the light of White's value-analysis technique (10) and Murray's analysis of major needs (6). (The value items used are given in Table 1.) In addition to these 35 items, the pack of cards presented to each S also contained value terms that had been coded out of his questionnaire answer to a verbal probe in conjunction with the item on "allowing members of the Communist Party to address the public." For most Ss the total number of such "salient" values came to two or three, the range among all Ss extending from zero to six.

For the "value importance" measure S was asked to rank each card so as to indicate how much satisfaction he gets, or would get, from the value state that it described. Each value was to be judged independently in terms of 21 categories ranging from "gives me maximum satisfaction" (Category +10) through "gives me neither satisfaction nor dissatisfaction" (Category 0) to "gives me maximum dissatisfaction" (Category -10). For the measure of "perceived instrumentality" he was asked to judge and place each card in terms of 11 categories ranging from "the condition (value described on the card) is completely attained by allowing admitted Communists to address the public" (Category +5) through "whether or not admitted Communists are allowed to address the public is completely irrelevant to the attainment of the condition" (Category 0) to "the condition is completely blocked by allowing admitted Communists to address the public" (Category -5). In each of the two card-sorting tasks S was required, after the initial categorization was completed, to rank the cards *within* each category in terms of "value importance" and "perceived instrumentality" respectively.

The test-retest reliability of these two measures was studied with a subpopulation of 12 Ss, the second administration following the first by from four to five weeks. For each of these Ss Spearman's rho was computed as an estimate of the degree of correlation between the ranks assigned to the separate values on the first and second administrations respectively. For the measure of value importance, the median rho was .89 ($p < .01$); for that of perceived instrumentality, it was .74 ($p < .01$).

From the data obtained through the two card-sorting tasks ten indices were computed to permit testing the major hypotheses. Operational definitions of these indices are given in Tables 2-4.

RESULTS AND DISCUSSION²

Each hypothesis was tested by chi square,³ computed from a 3×4 table. In each such

² All of the results reported in this section have been successfully replicated with regard to the attitude area of "allowing Negroes to move into white neighborhoods." The replication was based on an analysis exactly parallel to the one described in the present report and upon identical measurement operations.

³ This technique was used rather than correlational methods because it was felt that the assumptions required by correlational methods (particularly the

table the four groups in terms of attitude position (extremely favorable, moderately favorable, moderately unfavorable, and extremely unfavorable toward "allowing members of the Communist Party to address the public") were cross classified in terms of a threefold categorization of the *Ss* on one of the indices derived from the card-sorting value measures.

Hypothesis 1

All of the four indices used to test Hypothesis 1 provide, as the hypothesis requires, for representation of both the importance to the individual of some set of values and his perceptions as to how these values are affected (with regard to their attainment or blocking) through the instrumental agency of the attitude object (the policy of "allowing members of the Communist Party to address the public"). By algebraically summing the importance-instrumentality products for each of the values it is possible to represent appropriately the interaction between positive and negative values on the one hand and perceptions of positive and negative instrumentality on the other. Thus a positive value (rated say as +7 in its capacity to satisfy the *S*) may be perceived as being attained through the policy of "allowing members of the Communist Party to address the public" (to the extent say, of a rating of +3 on "instrumentality"). In this case the product would be +21. If on the other hand, the same value were rated as being *blocked* by the policy of "allowing members of the Communist Party to address the public" (for example a rating of -3 on "instrumentality") the resultant product would be -21. Similarly, a value rated as yielding dissatisfaction and thus bearing a minus sign on "value importance" would yield a negative product if multiplied by a rating of positive instrumentality and a positive product if multiplied by a rating of negative instrumentality. By summing all such products for a given set of values a single algebraic quantity is obtained representing the total import of the *S's* pattern of beliefs about the influence of the attitude object upon

assumptions of normality and equal-interval scales) were not fully met by the data. The method used for computing chi square was the "maximum likelihood test" of Mood (5) which is applicable to data plots with low cell counts.

TABLE 2
RELATIONSHIPS BETWEEN "VALUE IMPORTANCE
× PERCEIVED INSTRUMENTALITY" (INDICES
1-4) AND ATTITUDE POSITION
(*N* = 117)

Index	Description of Index	Chi Square	Probability*
1	Algebraic sum of importance-instrumentality products for all value items	26.33	.001
2	Algebraic sum of importance-instrumentality products for the twenty values ranking highest on importance	30.82	.001
3	Algebraic sum of importance-instrumentality products for all "salient" values	45.09	.001
4	Algebraic sum of importance-instrumentality products for all "non-salient" values	15.10	.02

* Examination of the actual data plots reveals that all of these significant chi squares are due to positive, monotonic relationships.

attaining or blocking various states that he values to differing degrees.

Since all four of the indices used to test Hypothesis 1 were found to be significantly related to attitude position (see Table 2), the hypothesis would appear to have been confirmed at an acceptable level. These findings, to generalize, lend support to the view that beliefs associated with an attitudinal affect tend to be congruent with it; i.e., that there exists within the individual an "organization" of the affective and cognitive properties of his total pattern of response to what is, for him, an "attitude object."

One possible objection to such an interpretation might hold that the way in which *S* sorts the cards may reflect not his actual cognitive structure but rather an attempt to demonstrate to the experimenter and to himself that his beliefs are consistent with his attitudes. If striving for cognitive-affective consistency accounts in part for relationships obtained during the testing session, however, it seems appropriate to assume that similar motivation had been in operation before the testing session, i.e., that the motive to "rationalize" one's attitudes plays a role in the natural history of attitude development.

Still another line of evidence that limits the significance of this objection arises from the findings obtained with Index 3. This index is based solely upon the "value importance" and "perceived instrumentality" ratings of the *S's*

"salient" values, which were coded out of his own verbal defense of his stated attitude position. Certainly the individual's estimates of the importance of these values and the ways in which their realization would be influenced by the attitude object must have had *pre-experimental* reality to have been given expression in response to the questionnaire probe.

It should be noted, however, that the historical relationship between affective and cognitive processes in the development of attitudes is not approachable through analysis of the present data. The province of the present study is to examine the relationships between cognitive and affective processes at a point in time after an attitude has been acquired and stabilized. Nevertheless, the present hypotheses may, if confirmed, suggest further hypotheses about the processes of attitude learning and attitude change.

Hypotheses 2 and 3

While the data just reviewed confirm the existence of a relationship between a measure of affect toward an object and a measure based upon both the "importance" and "perceived instrumentality" of values associated with it, the question may be raised as to whether each of these two latter variables may be demonstrated to covary with affect when the other is held constant. Hypotheses 2 and 3 assert the existence of such relationships. Indices 5-10 are based upon procedures for obtaining separate estimates of the "value importance" and "perceived instrumentality" variables. To isolate the one variable, it was possible to hold the other constant since the original card-sorting tasks had required *S* to rank the values *within* each of the categories used in sorting.

Thus, to test the relationship between "perceived instrumentality" of values and affect toward the attitude object, four indices (5-8) were developed, all of which have in common being based upon a summation of the "perceived instrumentality" ratings of a group of values defined and chosen in terms of their "value importance" ranks. Index 5 for example is obtained by choosing the value items ranked 1-20 by the *S* on the dimension of "value importance" and then obtaining the algebraic sum of the "perceived instrumentality" ratings of these same 20 values.

A similar procedure was employed with regard to the indices (9 and 10) used to test the

relationship between "value importance" and affect toward the attitude object. These two indices are both based upon the "value importance" ratings of a group of values defined and chosen in terms of their "perceived instrumentality" ranks. Index 9 for example is obtained by choosing the five values ranked highest on "perceived positive instrumentality" and then obtaining the algebraic sum of the "importance" ratings of these same five values. Index 10 is identical except that it is computed from the five values ranked highest on "perceived negative instrumentality."

The results reported in Table 3 indicate that all four of the indices of "perceived instrumentality" are significantly related to attitude position. Furthermore it should be noted that the hypothesis has been verified at a number of different levels of "value importance" control, each of the four indices representing a different level. On the basis of these findings Hypothesis 2 appears to have been confirmed. To generalize, other things (particularly "value importance") being equal, extreme attitudinal affects are associated with perceptions of close positive or negative instrumental connection between the attitude object and related values, while moderate attitudinal affects are associated with perceptions of less clear-cut instrumental relationships. In terms of attitude theory this finding suggests the existence of a distinguishable and important

TABLE 3
RELATIONSHIPS BETWEEN "PERCEIVED
INSTRUMENTALITY" (INDICES 5-8) AND
ATTITUDE POSITION (*N* = 117)

Index	Description of Index	Chi Square	Probability*
5	Algebraic sum of the instrumentality ratings of the values ranking 1st-20th on importance	35.51	.001
6	Algebraic sum of the instrumentality ratings of the values ranking 11th-30th on importance	21.27	.001
7	Algebraic sum of the instrumentality ratings of the values ranking 1st-10th on importance	20.82	.01
8	Algebraic sum of the instrumentality ratings of the values ranking 11th-20th on importance	38.47	.001

* Examination of the actual plots reveals that all of the significant chi squares are due to the presence of positive, monotonic relationships.

TABLE 4
RELATIONSHIPS BETWEEN "VALUE IMPORTANCE"
(INDICES 9 AND 10) AND ATTITUDE
POSITION ($N = 117$)

Index	Description of Index	Chi Square	Probability*
9	Algebraic sum of the importance ratings of the values ranking 1st-5th on positive instrumentality (attainment of values)	12.75	.05
10	Algebraic sum of the importance ratings of the values ranking 1st-5th on negative instrumentality (blocking of values)	16.46	.02

* Examination of the actual plots reveals that the significant chi square in the case of Index 9 is due to the presence of a positive monotonic relationship, while the significant chi square in the case of Index 10 is due to the presence of a negative monotonic relationship, both as predicted.

dimension of attitude-related cognitive structures. Recent research (1) based upon the present study lends confirmation to the existence of this dimension as a separate and manipulable one and indicates that attitude-change effects (in the sense of affective change) may be achieved through its manipulation.

The results reported in Table 4 indicate that both of the indices of "value importance" are significantly related to attitude position. On the basis of these findings, Hypothesis 3 appears to have been confirmed.

Although the relationships reported in Table 4 reach an acceptable level of significance, however, their significance level is lower than most of the relationships reported in Tables 2 and 3. The present data leave it an open question as to whether this is an artifact of the measurements or computations employed or whether, on the other hand, "perceived instrumentality" actually controls more variance in attitudinal affect than does "value importance." The generalization is nevertheless supported that other things (particularly "perceived instrumentality") being equal, extreme attitudinal affects are associated with values of high importance while moderate attitudinal affects are associated with values of less importance. For attitude theory, these results suggest the existence of a second distinguishable and important dimension of attitude-related cognitive structures. Further research is needed to determine whether the individual's assessment of the importance of his values is manipulable and whether, if it is,

such manipulation is sufficient to produce change in affect toward an object seen as having instrumental influence upon the attainment or blocking of these values.

Further Implications for Attitude Theory

It has been suggested that attitudinal affect toward an object may be altered by the prior modification of value importance and perceived instrumentality. A related implication is that one way in which attitudes may originally develop is through the prior acquisition of beliefs about the value-attaining or value-blocking powers of particular objects, be those objects individuals, groups, political proposals, or commercial products. Such a view, even if it has not been systematically elaborated, seems to underlie the attitude-creating and attitude-changing communication techniques of many propagandists in such fields as advertising, teaching, psychological warfare, and child-rearing.

A less obvious implication of the present findings suggests at least one other type of sequential process leading to the establishment of stabilized affective-cognitive patterns. It seems plausible that much original attitudinal learning (the acquisition of a stable affective response where none existed previously) may originate in experiences of being rewarded or punished for imitation or rehearsal of expressions of affect provided by others. A case in point may be found in the data on the acquisition of the anti-Negro attitude as reported by Horowitz (4). With an affect, or its beginnings, established through reinforcement procedures, something like a need for affective-cognitive consistency may set the person to acquire socially available beliefs that "rationalize" the acquired affect. When such beliefs are unavailable in the person's communicative surround, he may invent them. In real-life situations where attitudes are being instilled, there are generally available both cognitive supports for the advocated affect and direct reinforcements for its expression.

Recent research by the investigator seems to indicate that the production of *change* in an already established attitude may be obtained without any direct attempt to modify the associated cognitive structure, but rather through direct assault (with hypnotic techniques) upon the established affect. Such assault produces in some Ss a temporary affective reversal, and

along with this reversal, a spontaneous, self-directed modification of the related cognitive structure.

SUMMARY

An attitude questionnaire on the issue of "allowing members of the Communist Party to address the public" was administered to 117 Ss. Three to five weeks later each S took a card-sorting test which required him to rate and rank each of a group of value items in terms of (a) the importance of the value as a source of satisfaction and (b) his perception as to the extent to which the value tends to be attained or blocked through the instrumental agency of the attitude object.

The data thus gathered were used in testing three hypotheses concerned with the relationships between stable affective response (attitude) toward an object and beliefs about that object. These hypotheses were derived from a general proposition to the effect that the sign and extremity of affect toward an object are functions both of whether it is perceived as facilitating or blocking the attainment of values and of whether or not the values involved are important ones. All three hypotheses were confirmed. Some of the data were interpreted as suggesting that "value importance" and "perceived instrumentality" are separate and possibly, manipulable dimensions of attitude-related cognitive structures.

Some suggestions were presented as to the implications of the present study with regard to the phenomena of attitude learning and attitude change.

REFERENCES

1. CARLSON, E. R. Attitude change and attitude structure. *J. abnorm. soc. Psychol.*, 1956, **52**, 256-261.
2. CARTWRIGHT, D. Some principles of mass persuasion. *Hum. Relat.*, 1949, **2**, 253-268.
3. HILLIARD, A. E. *The forms of value*, New York: Columbia Univer. Press, 1950.
4. HOROWITZ, E. L. The development of attitude toward the Negro. *Arch. Psychol.*, N. Y., 1936, No. 194.
5. MOOD, A. E. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.
6. MURRAY, H. A. *Explorations in personality*. New York, Oxford Univer. Press, 1938.
7. ROSENBERG, M. J. The experimental investigation of a value theory of attitude structure. Unpublished doctor's dissertation, Univer. of Michigan, 1953.
8. SMITH, M. B. Personal values as determinants of a political attitude, *J. Psychol.*, 1949, **28**, 477-486.
9. TOLMAN, E. C. A psychological model, in T. Parsons & E. A. Shils, (Eds.), *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1951.
10. WHITE, R. K. *Value analysis: the nature and use of the method*. Society for the Psychological Study of Social Issues, 1951.
11. WOODRUFF, A. D. Personal values and the direction of behavior. *Sch. Rev.*, 1942, **50**, 32-42.
12. WOODRUFF, A. D., & DiVESTA, F. J. The relationship between values, concepts, and attitudes. *Educ. psychol. Measmt.*, 1948, **8**, 645-660.

Received January 23, 1956.

THE EFFECTS OF EMOTIONAL AROUSAL ON THE RETENTION OF FILM CONTENT: A FAILURE TO REPLICATE

ELEANOR E. MACCOBY, HARRY LEVIN,¹ AND BRUCE M. SELYA

Harvard University

IN A preliminary report (2) we described a study in which groups of children were subject to experimental variation in their emotional states before being shown a movie. The nature of the movie content recalled by the children a week later was reported to be a function of their emotional state at the time of viewing. In an attempt to replicate the study, however, we have since obtained negative results. Since the original abstract has been rather widely cited,² it seems desirable to publish these negative findings.

The hypothesis tested in the two studies was: if children are frustrated just before they see a movie, so that aggressive impulses are aroused, they remember more of the aggressive content of the film than do children who are not frustrated. This hypothesis derives from the assumption that perception is to some degree selective, so that subjects who are aggressively aroused perceive more of, or attend more vigilantly to, the aspects of the material that are congruent with their drive state.

An alternative formulation, that derives the same hypothesis from other premises, concerns the drive-reducing function of fantasy. The assumption that wish-fulfillment fantasies are satisfying is shared by many personality theorists, and Feshbach (1) has demonstrated that at least some kinds of fantasy may reduce an experimentally aroused drive. If under conditions of frustration, aggressive content is more drive reducing than other content, the present hypothesis could be derived in terms of either of two possible mediating processes: (a) The aggressive actions of fantasy heroes, which presumably provide some drive reduction to the viewers, serve as stimuli that may form habits with some implicit, probably verbal responses, so that under recall conditions their recall probability is higher than for those fantasy stimuli that are not associ-

ated with drive reduction; or (b) Possibly the drive-reducing aspects of the movie acquire thereby a pleasant affective tone, thus increasing the chances of spontaneous practice of the material during the interval before the recall test.

The following report describes the *second* in the series of two experiments.

METHOD

Subjects

The experiment was conducted in two semirural schools in upper New York state. The subjects were the children in eight fifth- and sixth-grade classrooms, numbering 190 children in all.

Procedure

Frustration. The first experimenter, a man, was introduced to the class as an expert on spelling. He told the children that there was to be a statewide spelling contest, that he was about to run a spelling bee in their class, and if the winning team were good enough, they might have a chance to enter the contest. The class was divided into two teams, the "reds" and the "blues," and the members were given appropriately colored ribbon badges. The two teams in each classroom were matched for sex and spelling proficiency as rated by the teachers, and a random procedure was used to determine which team was red, which blue. During the spelling bee, each child was allowed two mistakes before he had to sit down. The red team was given ninth-grade words to spell, and the blue team got fourth-grade words. The experimenter was cold, critical, and sarcastic toward the red team and friendly and supportive toward the blue team. The red teams were spelled down quickly, and began to complain of unfair treatment. The experimenter asked them if they thought they could do better if they had a second chance. They were then permitted to repeat the spelling bee, with the blue team continuing to receive favored treatment. At the direction of the experimenter, a member of the winning team collected the badges of the losing team, while the winning blues were permitted to keep theirs. (The fact that the children were highly involved in the contest is suggested by the fact that one little girl on a winning team got up early every morning for a week to iron her badge so she could wear it to school!) In some classrooms, the meaning of the colors was reversed, so that the reds were favored over the blues. After the badges of the losing team had been collected, the first experimenter left the room.

Exposure to the film. Toward the end of the spelling bee the second experimenter (a woman) came into the room and sat at the back of the room. After the spelling

¹ Now at Cornell University.

² See for example the report of the hearings of the Kefauver Committee on Juvenile Delinquency, specifically the hearings on the effect of television programs, April 6 and 7, 1955.

bee, she came forward and apologized for coming on such a busy day. (Every effort was made to dissociate the spelling bee and the movie.) She told the children that she was interested in finding out what sorts of things children liked and remembered in movies, and said she had a movie for them to look at, mentioning that she might ask them a few questions about it afterwards. The first chapter of a serial movie was then presented, starring the Dead End Kids and the Junior G Men. At the end of the movie which involved considerable aggressive action, E-2 told the children that there was no time for discussion now, and promised to come back in a week with another chapter of the film.

Measures for memory of movie content. A week later, the children were given a 66-item multiple-choice recognition test on the content of the movie they had seen. Items were of the following four kinds of items: (a) Seventeen aggressive central items, pertaining to the occurrence of overt aggressive acts, the identity of the aggressor, or the identity of the recipient of aggression. (Example: "In the fight between Billy's gang and the Junior G Men, who struck the first blow?") (b) Twelve aggressive incidental items, concerned with details surrounding aggressive acts. (Example: "In the fight that takes place immediately after the two trucks smash, where do the people watching the fight stand?") (c) Eighteen neutral central items pertaining to content that was central to the theme of the story, but did not involve aggressive action. (Example: "Why was Billy's father important?" the answer being that he had invented a new kind of explosive.) (d) Seventeen neutral incidental items concerned with content related to nonaggressive material incidental to the main story theme. (Example: "Jim Bradford had a very pretty secretary in his office. What was her name?")

Following the test, which took about 40 minutes, the children were shown the final chapter of the serial and were given an explanation of the experiment.

RESULTS

There were no significant differences between the frustrated and nonfrustrated groups in their recall of any of the different categories of film content. The team, rather than the individual child, was taken as the unit of analysis. Contrary to the hypothesis, there was a slight tendency for the nonfrustrated teams to remember more aggressive central content, but while six classrooms showed a difference in this direction, two showed a difference in the opposite direction. Neither a sign test, a *t* test, nor analysis of variance revealed a significant contribution of the experimental treatment to the children's recall of central aggressive content. Similarly, the frustrated and nonfrustrated subjects did not differ in their total recall scores, nor in the recall of aggressive incidental, neutral central, or neutral incidental content.

COMPARISON OF THE TWO EXPERIMENTS

The procedure for the two experiments was as nearly identical as possible. The identity of the two experimenters was different, and in the first experiment both experimenters were men. Our impression is that in both experiments, the experimental variable "took," in the sense that the children seemed much involved in the spelling bee and were anxious to win.

In the first experiment, the subjects were children living in a suburban section of the Boston area, while in the second, they were rural and small-town children in New York state. It is possible that the latter group have been less permissively treated with respect to aggression. And possibly, selective recall is a function not only of the amount of aggressive motivation present at a given time, but also of the child's level of aggression anxiety (growing out of past punishment for aggression). Presumably, if aggression anxiety is great enough, it leads to repression of aggression-related content, at least among some children. If the children in the second experiment had been more severely socialized with respect to aggression, then, this fact might account for their failure to remember more aggressive content under conditions of aggression arousal. We cannot test this hypothesis directly, since we have no information concerning the severity of socialization of aggression in these two groups of children. An indirect test is available, however. If the frustrated children in our second experiment were subject to a fairly high level of aggression anxiety, these children should be more *variable* in their recall performances than their nonfrustrated counterparts. Examination of the data shows that this was not the case. Therefore, we have no evidence for the hypothesis that the two sets of subjects differed in their level of aggression anxiety, although the possibility still exists.

REFERENCES

1. FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, 50, 3-12.
2. MACCOBY, ELEANOR E., LEVIN, H., & SELYA, B. M. The effect of emotional arousal on the retention of aggressive and nonaggressive movie content. *Amer. Psychologist*, 1955, 10, 359. (Abstract)

Received June 25, 1956. Prior publication.

CASE REPORT

THE RESPONSE OF SCHIZOPHRENIC PATIENTS TO A TELEVISED WORLD SERIES GAME: A STUDY IN SOCIAL ISOLATION

JEROME L. SINGER AND HERBERT E. SPOHN¹

Franklin D. Roosevelt V. A. Hospital, Montrose, N. Y.

PERHAPS the most typical of all the manifestations of a chronic schizophrenic state is the social isolation, the withdrawal from even the most rudimentary aspects of common social interchange. Individual interviews with patients from a schizophrenic ward may yield considerable heterogeneity in the appearance of psychotic symptomatology and personality characteristics, ranging from florid hallucinatory or delusional trends through monosyllabic response or silliness. Observation of these same patients as a social group in a ward dayroom or in various group settings suggests considerably more homogeneity, however, in the relative absence of what might be considered normal gregariousness or group behavior. Even in hospital settings (such as the one in which this study was carried out), where the personnel-patient ratio is not too unfavorable and where a full schedule of therapeutic and recreational activities is in effect, the absence of social intercourse, of mutually shared activities of a spontaneous sort, is striking. A ward dayroom of forty patients viewed on a typical afternoon following return from activities might reveal perhaps seven or eight men in conversation or engaging in a game of cards, while the rest of the group is scattered around the room with practically no sign of social interchange. Some patients may be moving restlessly, others talking in autistic fashion, others reading, others in catatonic posture, and some sleeping. One need only think by contrast of the hubbub that characterizes the dayroom of an army barracks or the courtyard of a prison to sense the dramatic depth of the social isolation of the schizophrenic. The brief study to be described here represents an initial effort to demonstrate this social isolation in a fairly objective manner and to test

certain hypotheses relating social isolation to chronicity and severity of illness. As part of a long-term project dealing with the social interaction patterns of chronic hospitalized psychotics, it represents one attempt to evaluate the response-potential of schizophrenics in various group situations.

The use of reaction to a televised World Series game was dictated by the well-known response of normal males in the same age and cultural groupings as the patients to such a stimulus. Whether the audience is in the ball park or in front of a television screen, the excitement engendered by a Yankee-Dodger series quickly lowers barriers to communication among strangers. One can delineate responses to the game, e.g., cheering, hooting, or crying encouragement at players, etc., or responses to others in the crowd, betting, praising, or damning the others' favorite, or arguing the merits of a player. The questions posed in this study were: (a) Would the normal enhancement of interaction and social responsiveness characteristic of World Series viewing be observable in schizophrenics? and (b) Would it be possible to demonstrate a progression in interaction from the most severely disturbed chronic wards through the less chronic and open wards, the latter group showing a response more like that of a non-hospitalized population?

PROCEDURE

Employing a standardized rating sheet and guide, eight clinical psychologists who were stationed on eight previously-selected wards of a large mental hospital observed the response of patients during three innings of the televised third game of the New York Yankees-Brooklyn Dodgers World Series in 1955. Since the Yankees had taken a two-game lead in the series, this game might be expected to be most important. The eight wards selected were Ward A, a chronic neurological ward in which most of the patients were not psychotic and hence might serve to some extent as a control group of chronic hospitalized individuals; Ward B, the open ward of the continued treatment service, the last stop before discharge, as a rule; Ward C, the acute

¹ The authors wish to express their indebtedness to Mrs. Mildred Rosenberg, Drs. Harold Wilensky, Frank Wexler, Ralph Colvin, Frank Lachmann, and to Mr. Frank Kennedy, who served as observers for the study.

TABLE 1

PERCENTAGES OF PATIENTS AND PERSONNEL WATCHING AND INTERACTING DURING TELEVIEWED GAME

Ward	Percentage Watching		Percentage Interacting		Tension Level	
	Pa-tients	Per-sonnel	Pa-tients	Per-sonnel	Patients	Personnel
A	90	100	49	100	Moderate	Moderate
B	90	100	47	100	High	Moderate
C	59	100	10	100	Moderate	Moderate
D	27	100	20	80	Moderate	High
E	34	100	0	100	Low	Moderate
F	31	100	20	100	Low	High
G	24	100	0	67	Low	Moderate

service ward for new admissions, a group generally younger than the hospital mean and characterized by recent onset and relatively florid symptomatology; Ward D presumably the most advanced of the continued treatment service wards with a greater number of patients participating in hospital industrial therapy activities; Wards E and F two wards of chronic patients of approximately equal severity of over-all pathology; Ward G, a disturbed ward of chronic patients who were occasionally assaultive or elopement threats; and Ward H, a ward of elderly chronic schizophrenic patients, many of them incontinent. It was expected that the response to the game in terms of numbers watching, overt reaction in terms of cheering, etc., and interaction among viewers would decline systematically from Ward A to H. The rating sheet called for the observer to record data during the first, fifth, and ninth innings on the number of patients and staff present (the TV sets were located in ward dayrooms), the number actually *watching* the game (rather than occasionally staring up at the set), the seating arrangement, the amount of interaction between patients and between personnel, the appropriateness of response (cheering during exciting plays as against cheering during a lull for a team not in the game), the rate of tension and excitement during suspenseful moments among patients and hospital personnel, and the various modes of interaction among the viewers. From these observations it was possible to calculate data on the ratio of actual watchers to potential audience and the amount of social interaction within the audience. It should be noted that each ward contained approximately forty patients and that means for age and chronicity of illness were approximately the same for all wards except Ward C. The neurological control group of Ward A were slightly older but with similar duration of hospitalization as Wards B, D, E, F, and G.

RESULTS

It proved possible to obtain adequate ratings for seven of the eight wards in question. On Ward H the aides reported that since no interest in the game could be elicited, they had tuned the set to a cowboy movie, which was not, however, eliciting any apparent response from the patients at the time of observation.

Results for the other seven wards are presented in Table 1. Data on spontaneous seating arrangements were not available since on all wards the aides arranged most of the chairs in rows in front of the TV sets while the patients were at lunch and, hence, more or less enforced a seating arrangement.

It is apparent from consideration of Table 1 that on the whole the game drew an unusually small audience. The percentage of those watching falls roughly into three groups, the neurological and open-ward patients showing a high viewing percentage, the young, acute schizophrenics, a moderate response, and the chronic schizophrenics generally undifferentiated in showing a very low response. Differences within a ward for the three innings were so slight as to warrant combination of the inning data. When the number of patients interacting with each other or the game was considered, the results are even more striking in demonstrating the absence of social response. The largest percentage of overt response emerged again from Wards A and B, but even here the percentages were low by comparison with the nearly 100 per cent overt response from the small number of attendants and other personnel who were watching the game in the dayrooms. Excitement as manifested verbally and motorically varied in similar fashion with relatively little evidence of reaction from the patients at moments which brought loud cheering from the ball-park audience and the attendants in the dayroom. No striking differences in appropriateness of response as a group occurred between the patient or attendant audiences, although on occasion a purely autistic reaction did come from a patient.

A number of qualitative observations may help to fill out the picture of severe isolation suggested by these results. Since the chairs had been oriented toward the TV set by the aides, the patients on returning from lunch automatically grouped themselves in relationship to the chairs. Thus, an initial glance into a ward suggested a much larger audience than turned out to be the case, since many patients who ordinarily stood next to chairs against the wall facing into the dayroom now oriented themselves toward the TV set. Closer observation revealed that only a few patients seated in front were actually watching the game. This impression was confirmed by inter-

viewing patients individually later in the game. Occasionally the picture and sound, perhaps because of their figural quality in the general configuration of the room, evoked an upward glance from patients who otherwise showed not the slightest awareness of the nature or course of the game. It should be stressed that many of these patients were able to verbalize appropriately on other topics in response to inquiry. Among the chronic patients who responded to the game on Wards D, E, F, and G, a high percentage turned out to be persons with borderline diagnoses of schizophrenia; they tended to be persons with "character disorder" or "alcoholism" as part of their classification and were less clearly schizophrenic in their other symptomatology.

A dramatic example of the contrast between patients and hospital personnel was the situation on Ward D. There were three barbers at work in the dayroom cutting hair during the game. They kept up a running patter of wisecracks, comments about various players' skill, and wagers about the outcome of the game. By contrast, of the dozen or so patients watching the game, only two spoke at any time or cheered appropriately during exciting moments. These two patients occasionally made comments to the others and received no response. The barbers on occasion exchanged comments with these two patients about the game, but when in their enthusiasm they made remarks to other patients or slapped them on the back, they elicited no response. When Campanella hit a home run, the barbers and the two patients shouted loudly, but there was no other cheering in the room. Apparently in response to the tension and shouting, certain patients who were obviously not looking at the TV set muttered to themselves and moved about restlessly or went through manneristic rituals. Some type of social communication passed through the patients on the ward, but at a decidedly rudimentary level of complexity. An example of this primitive social awareness was the way in which one patient, as another rose from the barber chair, walked silently

from a separate part of the large room to take his seat for a haircut.

DISCUSSION

The problem of understanding how human beings who are very much alive in terms of retention of intellectual functions and, in individual interviews, sensitivity to misunderstanding or rejection, can so completely withdraw their interest from others in social groups is highlighted by these findings. Lack of interest in sports was never a striking feature of the premorbid histories of these men as far as could be determined, and so it cannot account for the apathy and absence of interaction. Nor can it be asserted that these patients have lost themselves in daydreaming or private worlds of conscious reverie because interviews and psychological tests of fantasy-disposition suggest that chronic schizophrenic patients engage in relatively little internal meditation. Perhaps chronic anxiety and fear of attack or rejection, as well as the unsatisfying nature of fantasy, have led to a kind of freezing of all modes of expression. Possibly, if one feels that some energy outlet is essential, it may take the form of endless concentration on trivial or commonplace stimuli, e.g., watching the ash burn down on a cigarette tip, noting the gurglings of one's own digestive system, or studying a repetitive drapery design.

Whatever the manner in which the chronic schizophrenic manages to maintain his isolation, the *fact* of his immense alienation from social interchange is highlighted by this study. Attempts at establishment of a therapeutic milieu through recreational and related facilities in the treatment of large numbers of patients are obviously essential in maintaining a "normal" atmosphere for patients. Yet they may be expected to prove frustrating until such time as social or clinical research provides some comprehension of the manner in which human beings can so fully detach themselves from interpersonal communication in group situations.

Received February 1, 1956.

CRITIQUE AND NOTES

AUTHORITARIANISM AND INTERPERSONAL PERCEPTION¹

W. H. CROCKETT AND THOMAS MEIDINGER

Kansas State College

RECENTLY Scodel and Mussen (7) and Scodel and Freedman (6) have examined the relationship between authoritarianism, as measured by the California F scale, and accuracy of interpersonal perception. Their results indicate that individuals who make high scores on the F scale tend to estimate that their peers will also score high on the F scale, whether these peers are actually high or low. On the other hand, individuals who make low scores on the F scale are less uniform in their judgments, but tend to estimate peers as being either moderate or high in their responses to the F scale. No differences were found in the *accuracy* of estimates by high F and low F subjects. The present paper reports a replication of these studies, using a different population and a different operational definition of accuracy of judgment.

METHOD

Subjects (Ss) were drawn from three classes in general psychology at Kansas State College, and from the 80 residents of the men's dormitory. Students in general psychology were required to participate as part of their course work, while dormitory men were asked to volunteer. Of the 80 dormitory men asked, only one refused because of a lack of time.

The experimental design paralleled that used at Ohio State University by Scodel and Mussen and by Scodel and Freedman. Pairs of Ss were asked to discuss for 20 minutes a topic suggested by the experimenters. Each S had previously responded to the F scale. After the discussion Ss were asked to fill out the F scale as they felt their partners in the discussion would respond to it.

The Ss were assigned to pairs according to their free hours. No Ss were paired who lived at the same address or in the same section of the dormitory. Members of minority groups were excluded from the experiment. Ninety pairs of Ss were used. In 30 pairs the F-scale scores of both members were above the median of their sex; in 30 other pairs the scores of both members were below the median of their sex; and in the remaining 30 pairs one S scored above the median of his sex while the other scored below.

Although this study was patterned after the Ohio State studies, the following modifications in experimental design and analysis were introduced:

a. *Measurement of accuracy.* On the Ohio State

studies accuracy is operationally defined as the extent to which the S estimates his partner's *total score* on the F scale. Thus, if the partner's total F-scale score is 127, the S is termed "accurate" if his estimate of the partner is very close to 127 and "inaccurate" if his estimate deviates markedly from 127. It has recently been pointed out (3, 5) that such a definition of accuracy has severe limitations, since any given total score on a multi-item test (excepting the highest and lowest possible scores) may be achieved in many possible ways. Since there are empirical as well as theoretical grounds for the existence of several factors in the F scale (1, 2), it seems likely that personality differences may be reflected in the *pattern* of an S's responses to the individual items. An alternative definition of accuracy in interpersonal perception therefore seems desirable: the degree to which the S reproduces the pattern of his partner's responses. Accordingly, the profile-similarity test presented by Cronbach and Gleser (4) was used in this study as an additional measure of accuracy.

b. *Questionnaire used and scoring method.* The Ohio State studies employed a 28-item version of the F scale, while the present study utilized the 30-item scale presented by Adorno *et al.* (1). The method of scoring used in this study also differed from that used in the Ohio State studies. The F-scale scores in the Ohio State studies were obtained by adding a constant of 4 to the value which the S had assigned to each F item. Thus, a -1 response was scored 3, and responses of +1, +2, and +3 were scored 5, 6, and 7 respectively. In the present study, a -1 response was scored 3, as in the Ohio State studies, but responses of +1, +2, and +3 were scored 4, 5, and 6 respectively.

These differences in the items used and in the scoring of items probably would not alter the relative positions of respondents in their respective groups. However, the differences do rule out direct comparisons of the scores of Ss at Ohio State with those at Kansas State.

c. *Discussion topic.* In the Ohio State experiments all pairs of Ss were assigned "radio, television, and movies" as a discussion topic. The Ss in the present experiment were divided into two groups, one of which discussed radio, television, and movies while the other discussed child-training practices. Since the topic of discussion had no effect upon accuracy of judgment,² results from these groups will be combined in the present report.

d. *Range of respondents included.* All Ss in the Ohio State studies made either extremely high or extremely low scores on the F scale. The present experiment included as Ss moderate as well as extreme respondents. However, to increase comparability, we present only results based upon estimates by those Ss who scored in the highest and lowest quartiles.

¹ Portions of this paper were presented at the September, 1955, meetings of the American Psychological Association. Financial assistance in conducting this study was provided by the Faculty Research Fund of the School of Arts and Sciences, Kansas State College.

² This aspect of the experiment will be reported fully in a separate publication.

TABLE 1

ACTUAL AND ESTIMATED F-SCALE SCORES IN THE SCODEL-MUSSEN STUDY, THE SCODEL-FREEDMAN STUDY, AND THE PRESENT STUDY

Group	Actual Score of Judging Group		Estimate of Partner's Score		Diff.	<i>t</i>	<i>p</i>
	Mean	SD	Mean	SD			
High F Judging High F							
Scodel-Freedman (<i>N</i> = 24)	125.46	7.68	122.33	12.13	3.13	1.07	.20
Crockett-Meidinger (<i>N</i> = 29)	125.52	5.58	122.38	9.78	3.14	3.14	.01
High F Judging Low F							
Scodel-Mussen (<i>N</i> = 27)	129.26	9.34	124.19	11.08	5.07	1.79	.10
Crockett-Meidinger (<i>N</i> = 16)	123.25	5.04	123.81	10.01	.56	.46	.30
Low F Judging Low F							
Scodel-Freedman (<i>N</i> = 24)	67.04	11.94	99.75	18.06	32.71	7.42	.001
Crockett-Meidinger (<i>N</i> = 30)	83.13	10.71	99.10	29.56	15.97	6.46	.001
Low F Judging High F							
Scodel-Mussen (<i>N</i> = 27)	66.52	11.15	99.81	21.88	33.29	7.05	.001
Crockett-Meidinger (<i>N</i> = 14)	87.50	8.29	101.14	9.95	13.64	6.85	.001

TABLE 2

FREQUENCY OF ESTIMATED F-SCALE SCORES ACCORDING TO COMBINATION OF ACTUAL SCORES IN THE SCODEL-MUSSEN STUDY, THE SCODEL-FREEDMAN STUDY, AND THE PRESENT STUDY

Group	Estimated F-Scale Score		
	High*	Middle*	Low*
High Judging High			
Scodel-Freedman	16	8	0
Crockett-Meidinger	20	9	0
High Judging Low			
Scodel-Mussen	26	1	0
Crockett-Meidinger	12	4	0
Low Judging High			
Scodel-Mussen	9	13	5
Crockett-Meidinger	1	7	6
Low Judging Low			
Scodel-Freedman	6	14	4
Crockett-Meidinger	4	15	11

* In the Scodel-Mussen study Ss' estimates were called high if the total score was 119 or more, medium if the score was between 80 and 118, and low if it was 79 or below. In the Scodel-Freedman study the figures were high, 115 and over; medium, 80-114; and low, 79 and below. For the present study the figures for high, medium, and low categories respectively were: men, 117 and over, 97-116, and 96 and below; for women 116 and over, 96-115, and 95 and below. It should be pointed out again that the scoring technique used in the Ohio State studies was different from that used in the present study, and that the total scores are not directly comparable.

e. Sex differences in F-scale scores. Our experimental design called for one-third of the pairs in each condition to be women. The median F-scale score for men was five points higher than the median for women. If, as it seemed plausible to assume, this sex difference reflects a social norm for women considerably less "authoritarian" than that for men, assigning Ss to high F or low F status on the basis of the combined distribution would result in the misplacement of some Ss rela-

tive to the norm for their sex. Male and female Ss were therefore assigned to high F or low F status relative to the median of their own sex. Similarly, the distributions of profile-similarity scores were kept separately for men and women.

RESULTS

Table 1 presents for both of the Ohio State studies and for the present study the actual mean F-scale scores of high F and low F Ss and the mean scores that these Ss attributed to their partners.³ In all of these studies, the mean estimate by high F Ss was very similar to the mean of the high F Ss' own scores, regardless of the partners' actual positions on the F scale. In one comparison in the Kansas State study—that in which high F Ss judged high F partners—the mean of the estimates was significantly lower than the mean of the Ss' own scores. Nevertheless, the absolute magnitude of the difference between these means was quite small, especially when compared with the differences between estimated and actual means of low F Ss. For low F Ss, on the other hand, the mean of the estimates of all groups was significantly higher than the mean of the low F Ss' own scores, regardless of the partners' actual positions.

Table 2 further clarifies the divergent estimation behavior of high and low F Ss. This table presents the number of Ss in each group whose estimates of their partners fall within the high, middle, and low categories of the F-scale distribution. As may be seen, most high F Ss estimated their partners to be in the highest group, and no high F S attributed a low F-scale score to his partner. For low F Ss, on the other hand, much more variability is present

³ We are greatly indebted to Professor Alvin Scodel for making these data available for the Scodel-Freedman experiment.

in the low F estimates than in estimates by high F Ss.

The Ohio State experimenters have advanced as an explanation of this difference in variability between highs and lows a tendency for the authoritarians to project their own responses upon their partners. As Scodel and Mussen say, "The highs' strong need to identify with an ingroup of superior status results in perceptual distortions which exaggerate and extend the similarities between themselves and their peers" (7, p. 184).

To test this hypothesis we defined projection operationally as the similarity between the pattern of the S's own responses and the pattern of his estimates of his partner. The assumed similarity scores obtained were ordered in a frequency distribution for each sex. These distributions were divided at the median, and the number of high F and low F Ss falling above and below the median assumed similarity score was ascertained. High F Ss were no more likely than low F Ss to have high assumed similarity scores (chi square = 1.12, $p > .20$). Thus, the projection hypothesis was not confirmed.

As an additional test of accuracy of perception, for each S a profile-similarity score was computed comparing his estimate of his partner with the partner's actual responses. The distributions of these profile-similarity scores for each sex were dichotomized. High F and low F Ss did not differ significantly in the accuracy of their estimates as measured by this method (chi square = .12, $p > .70$).

DISCUSSION

It is evident that the results of the present study are strikingly similar to those at Ohio State. A number of explanations of the results are available. One possible explanation assumes that agreement with a number of F-scale items is characteristic of many fundamentalist and conservative social, religious, and political groups in our society. A high F S whose social experiences have been principally within such groups and who has

had little experience with vociferous proponents of contradictory views is likely to believe that a particular person will agree with him unless he has direct information to the contrary. Twenty-minute discussions probably do not provide much direct information about the partner's position on these items. On the other hand, in view of the wide dissemination given conservative, traditional views by our institutions of mass communication and by certain extremely vocal political groups, it would be unlikely for a low F S to avoid recognizing that a considerable number of people disagree with him about certain items of the F scale. Since the low F S's attitudes have not developed in a social vacuum, however he is also likely to be familiar with individuals and groups who share his own opinions. Thus, the overestimation of the degree of authoritarianism and the greater variability of low F Ss may simply reflect contemporary social conditions.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. CHRISTIE, R., & GARCIA, J. Subcultural variation in authoritarian personality. *J. abnorm. soc. Psychol.*, 1951, **46**, 457-469.
3. CRONBACH, L. J. Processes affecting scores on understanding others and assumed similarity. *Psychol. Bull.*, 1955, **52**, 177-193.
4. CRONBACH, L. J., & GLEESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456-473.
5. GAGE, H. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411-422.
6. SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, **52**, 92-95.
7. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, **48**, 181-184.

Received December 7, 1955.

ETHNIC CLEAVAGE AMONG YOUNG CHILDREN

WALLACE E. LAMBERT AND YOSH TAGUCHI

McGill University

AN APPARENT discrepancy exists between the age when preferential or prejudiced attitudes toward ethnically different people develop and the age when preferential or prejudiced behavior appears. A recent summary of research (6) places the appearance of prejudiced attitudes at the third or fourth years compared with the onset

of prejudiced behavior at the eighth to tenth years. Moreno (11) found no sociometric evidence of behavioral cleavage among young children of various ethnic backgrounds until age ten (Grade 5). Criswell (2), using the same technique, studied children from nursery school to the teen ages and found no cleavage until the eighth year (Grade

3). She noted that ethnic cleavage was most pronounced at the tenth year (Grade 5). The age determined by Moreno and Criswell is not only accepted as the established age of onset of prejudiced behavior in children, but their findings are also used as supporting evidence for theories of prejudice (e.g., 6, p. 1037; 8, p. 514).

The appropriateness of Moreno's and Criswell's method is open to question. Standard sociometric questions are often not meaningful to preschool Ss. In preliminary work we found that four- and five-year-olds generally "liked everyone" and "didn't mind" who sat next to them, suggesting that young children may not discriminate favored associates on such issues. Discriminations may be made, however, if the more significant values of preschool children are incorporated in the choosing procedures. To test this possibility, we measured ethnic cleavage among young children with techniques which, we assume, are more meaningful to them.

METHOD

The thirteen Ss of this study comprised the total membership of a small nursery school in Montreal. Six children were of Japanese origin (Oriental) and seven were of Roumanian, Greek, German, and Polish ancestries (Occidental), but all thirteen spoke English fluently. They attended school from 9:00 A.M. to 5:00 P.M. five days a week since both parents of all the children normally worked in Montreal. The newest member had been in attendance for over two months when the investigation was begun. None of the children in the school lived close enough to one another to play together outside of school. Two of the Oriental and three of the Occidental Ss were male. The ages for Oriental Ss were: 5.1, 5.6, 5.6, 5.8, 5.8, and 6.0; for the Occidental: 3.5, 4.0, 4.2, 5.5, 6.0, 6.0, and 6.2.

The experimenter (Taguchi, an Oriental) was introduced to the children as a guest supervisor (one regular supervisor was a white Canadian, the other an Oriental) and actively participated in the school program once each week for a month before the investigation commenced. All experimental work with the children began at 11:00 A.M. during the recreational period.

Type of choices. Four-year-olds generally place high value on receiving or giving candy and on having their pictures taken. It was assumed that choices of associates would be meaningful to children if these values were made part of the choosing procedure. The Ss were asked (a) to give away a piece of candy to an associate and then receive one for themselves (candy-giving choices); (b) to choose an associate to pose with them for picture taking (picture-taking choices); and (c) to indicate which picture, from among individual poses of each school member, they liked best and would want to keep, excluding their own (picture-choosing choices). In each of these tests the child was taken individually to a side room and given directions for each situation. The candy-giving test was repeated twelve times during a four-week period while the other two tests occurred once only.

TABLE 1
TEST CHOICES OF OCCIDENTAL AND ORIENTAL CHILDREN

	Candy-Giving Choices*		Picture-Taking Choices		Picture-Choosing Choices	
	In group	Out group	In group	Out group	In group	Out group
Occidentals	4	3	5	2	4	3
Oriental	6	0	6	0	6	0
Exact tests of significance		$p = .048†$		$p = .002†$		$p = .048†$

* Mean choices for twelve repetitions of the procedure.

† Two-tailed tests of significance.

RESULTS AND DISCUSSION

The choices made in the three test situations are presented in Table 1. Exact tests of significance (7, p. 303f) indicate a significant association between ethnic background and type of choices made, for all three situations. However, the tendency for Occidental children to choose within their own ethnic group is not a significant one since the probabilities of obtaining differences equal to or greater than a 4-3 and 5-2 split by chance alone are, respectively, 1.00 and .453 using an exact two-tailed test. Ethnic cleavage is clearly apparent, however, with the Oriental children. The Orientals' within-group preference is significant since the probabilities of obtaining a difference equal to or greater than a 6-0 split is .031 for each instance.

The findings indicate that ethnic cleavage appears among preschool children. We maintain that other researchers have not found prejudiced behavior at earlier ages because their methods of observation have not penetrated to the significant values of the younger Ss. Further research is clearly necessary in order to generalize this finding beyond Occidental-Oriental Ss and beyond so small and compact a group as that used here.

Davitz (3) has recently suggested that people have a basic need to be similar to valued persons, this need manifesting itself in a child's striving to be like his parents or an important peer. Davitz turns to Mowrer's theory of developmental identification (12) to account for such a need. For Mowrer, identification is equated with the secondary reward value of a person for the infant. Developmental identification starts with the interaction between the infant and the others in his social environment who satisfy his needs. When drive reduction and the cues associated with those who reduce the drives are repeatedly paired, the cues themselves take on secondary reward value for the infant. On a similarity gradient, school-

mates of the same ethnic background would be more alike, at least in physical characteristics, the family members with whom the child has identified. Any secondary reward generated during the identification process would be more likely to generalize to like schoolmates.

It is apparent that Mowrer's theory accounts for the behavior of our Oriental Ss but does not explain the Occidental group's behavior. Our finding of "one-way" cleavage is not unique. Many researchers have noted the preferential choice of within-group members with school-age children. Criswell (2) noted that, when racial cleavage did appear, it was the Negro Ss who were the first to segregate themselves. Radke *et al.* (13) found that Jewish children, 5-9 years old, were more emotionally identified with their own group than were Catholic or Protestant children. At the high-school age, Loomis (9), Goodnow and Tagiuri (5), and Lundberg (10) have all noted ethnic and religious within-group segregation.

It appears that minority group members (members of groups toward whom prejudice is commonly directed) are generally the first to segregate themselves from the majority group, this self-segregation occurring with preschool and young children as well as at the high-school level. Mowrer's and Davitz's theories can be extended to account for this empirical fact. Nurturance needs arising in infancy are satisfied by people who have distinguishable cue properties. The cues, acting alone as secondary reward agents, may become more active agents whenever nurturance-type needs are activated. Minority status, with its aspects of threat, would regenerate similar needs and concomitantly intensify the significance and vividness of cues associated with own-group members. The secondary reward accompanying the cue properties of within-group associates will reduce the nurturance needs of minority group members in a mixed-group situation. Majority group members, with less environmental threats, will not experience increased nurturance-type needs in a mixed-group situation and will accordingly receive less reward from within-group situations.

G. Allport has discussed the self-segregation phenomenon in other terms. He believes that within-group cohesion mitigates the problems of minority group members by assuring them friendship (1, p. 149), and permits them easier interaction where they need not learn new habits of social intercourse (1, p. 18).

A recent study by Goodman (4), however, poses

problems for both Mowrer and Allport, and suggests that further attention should be given to prejudiced behavior among young children. Goodman found that Negro preschool children often showed preference for *white* characters in stories, white dolls, and even white people.

SUMMARY

Ethnic cleavage among preschool children was studied using various test situations where choices of associates could be observed. It was found that ethnic cleavage does exist when choices are made significant to young children. The results are compared with other related findings and attempts are made to explain them theoretically.

REFERENCES

1. ALLPORT, G. W. *The nature of prejudice*. Cambridge, Mass.: Addison-Wesley, 1954.
2. CRISWELL, JOAN H. Racial cleavage in Negro-white groups. *Sociometry*, 1937, 1, 85-89.
3. DAVITZ, J. R. Social perception and sociometric choice of children. *J. abnorm. soc. Psychol.*, 1955, 50, 173-176.
4. GOODMAN, MARY ELLEN. *Race awareness in young children*. Cambridge, Mass.: Addison-Wesley, 1952.
5. GOODNOW, R. E., & TAGIURI, R. Religious ethnocentrism and its recognition among adolescent boys. *J. abnorm. soc. Psychol.*, 1952, 47, 316-320.
6. HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Ch. 27.
7. KENDALL, M. G. *The advanced theory of statistics*. London: Griffin, 1943.
8. KLINEBERG, O. *Social psychology*. New York: Holt, 1954.
9. LOOMIS, C. P. Ethnic cleavages in the southwest as reflected in two high schools. *Sociometry*, 1943, 6, 7-26.
10. LUNDBERG, A., & DICKSON, L. Selective association among ethnic groups in a high school population. *Am. sociol. Rev.*, 1952, 17, 23-34.
11. MORENO, J. L. *Who shall survive?* Washington: Nervous & Mental Disease Pub. Co., 1934.
12. MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
13. RADKE, M., TRAGER, H. G., & DAVIS, H. Social perceptions and attitudes of children. *Genet. psychol. Monogr.*, 1949, 40, 327-347.

Received December 1, 1955.

DIFFERENCES IN THE PERCEPTION OF AN AUTHORITY FIGURE AND A NONAUTHORITY FIGURE BY NAVY RECRUITS

WARREN J. WITTREICH¹ AND KEITH B. RADCLIFFE, JR.

Naval Medical Research Institute²

THE Navy recruit is introduced into military discipline immediately upon his arrival in boot camp. He is taught that orders given are orders obeyed—the alternative being severe disciplinary action. He rapidly learns to recognize those individuals who represent authority, and he becomes aware of the implications the these authority figures have for his own successful behavior.

In the early phases of training, when the recruit is still relatively unsure of himself, the appropriate visual configuration bearing stripes or bars is often sufficient to provoke fear and anxiety. As recruit training teaches him the appropriate forms of behavior in the appropriate situations, the element of fear is gradually diminished. Yet, seldom is that element completely eliminated. The comforting reliability of being in the presence of one's own peers is invariably preferred to the ever-present uncertainty involved in dealing with authority.

The question then arises: Does this element of fear involved in viewing and being with an authority figure actually influence the perception of that figure? Previous studies in this general area would indicate an affirmative answer. Gilder *et al.* (1) provide evidence which seems to indicate that a threatening figure changes less readily than a non-threatening figure under conditions of induced optical change. Wittreich and Radcliffe (2) have shown that the human figure under the condition of simulated mutilation is less susceptible to induced optical change than the human figure under conditions of normal configuration.

Hence the hypothesis is advanced that under conditions of induced aniseikonic distortion a figure that represents authority is more resistant to perceptual change than a figure that does not represent authority. This hypothesis was tested in a study of Navy recruits.

METHOD

Twenty-four white male Navy recruits in their seventh week of training at the Naval Training Center at Bainbridge, Maryland, served as Ss.³ The S was seated approximately 10 feet from a black backdrop.

¹ Now with the Armstrong Cork Company, Lancaster, Pa.

² The opinions or assertions contained in this article are the private ones of the authors and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

³ The authors wish to extend their sincerest thanks to Captain S. P. Sanford, MC, USN, and to Lt. W. B.

During the experimental session the S was able to view only a single figure standing before the backdrop. Illumination was provided by four 60-watt bulbs mounted in reflectors which were focused on the figure being observed.

Two figures were observed by each S: (a) *Authority figure*. The observed individual was dressed in a white enlisted man's uniform. On his left sleeve was the rating badge of a 1st Class Boatswain's Mate and two hash marks. (It should be pointed out that the Ss used had 1st Class Petty Officers as their Company Commanders.) (b) *Nonauthority figure*. The observed individual was dressed in a white enlisted man's uniform with the identifying marks of a recruit: canvas leggings or "boots" and the stripes of a Seaman Apprentice. (At Bainbridge all Seaman recruits wear the stripes of a Seaman Apprentice.)

Two different individuals played the roles of the authority and the nonauthority figure. Every effort was made to present these figures as genuine. Within each subsample of 12, one individual played the role of the authority figure 6 times and the nonauthority figure 6 times. The names assigned to the authority figure and to the nonauthority figure remained constant. The two individuals employed were approximately the same size, weight, and body build.

Each figure was viewed by the S through aniseikonic lenses in a series of 14 lenses ranging in power from 0.25 per cent to 4.50 per cent. Four series of lens presentations were given for each figure; 2 of the series were ascending, 2 were descending. The S attended a single experimental session which lasted approximately 50 minutes. The observed individual was visible to the S only during the actual presentation of each single lens power. After a lens had been inserted into the viewing apparatus, the S pressed a button which illuminated the figure standing in front of the backdrop. He then observed that figure for a period of 25 seconds, at which time an electronic timer shut off the light. Following each lens presentation the S reported the appearance of the observed individual for that particular lens presentation.

The E gave the following specific instructions: "Each time the light goes on you are to observe the appearance of the individual in the room as closely as possible. After the light goes off, I want you to tell me what he looked like when the light was on. If he looks the same as usual, I want you to tell me that. If he appeared to be changed in any way whatsoever, I want you to tell me that. I also want you to tell me in what way he changed. Now he may change in a number of ways. He may change in size; he may appear to tilt forward or backwards; he may change in his distance from you; he may change in specific parts of his body. Any one

Lyon, MSC, USNR, whose cooperation made this study possible.

or a combination of these changes may occur. In any event, please tell me after each lens presentation what his appearance was while the light was on."

The *E* recorded the point on the optical scale (lenses numbered 1 through 14) at which distortion of the observed individual was first reported when the series was ascending, or the point at which distortion was no longer reported if the series was descending. The starting point for any particular ascending or descending series was varied within conditions, but equated between conditions. Also, the order of presentation of the authority and the nonauthority figure, as well as the order of presentation of the ascending and descending series, was balanced within each subsample of 12 *Ss* so as to cancel out any possible effects due to order of presentation.

RESULTS

For the total sample of 24 *Ss*, the mean threshold for distortion for the Boatswain's Mate was 3.28 (*SD* 2.29); for the Seaman Apprentice 2.16 (*SD* 1.32). For 22 of the 24 *Ss*, the threshold for distortion was higher for Boatswain's Mate than

Seaman Apprentice. By sign test, this gives a two-tailed *p* value of less than .0001.

SUMMARY

As predicted from previous work on fear-evoking objects and resistance to perceptual distortion, Navy recruits viewing persons through aniseikonic lenses showed higher distortion thresholds when viewing an authority figure than when viewing a nonauthority figure.

REFERENCES

1. GILDER, R., THOMPSON, S. V., SLACK, C. W., & RADCLIFFE, K. B., JR. Amputation, body image, and perceptual distortion. *USN Med. Res. Inst. Rep.*, 1955. (Project NM 004 008.04.03.)
2. WITTEICH, W. J., & RADCLIFFE, K. B., JR. The influence of simulated mutilation upon the perception of the human figure. *USN Med. Res. Inst. Rep.*, June 1954. (Project NM 004 008.04.02.)

Received November 16, 1955.

A NOTE ON THE SOCIAL PERCEPTIONS OF AUTHORITARIANS AND NONAUTHORITARIANS

WILLIAM RABINOWITZ

Bank Street College of Education

SCODEL and Mussen (5) recently demonstrated that when authoritarian and nonauthoritarian persons, as measured by the *F* scale (1), were paired in a social situation, after which they estimated each others' *F*-scale responses, the nonauthoritarians were more accurate than the authoritarians. They interpreted their results as evidence that nonauthoritarians possess greater interpersonal sensitivity and insight than authoritarians. However, a more recent study by Scodel and Freedman (4) does not support this interpretation.

The Scodel-Freedman study essentially repeated the Scodel-Mussen design with one important change: authoritarians were paired with other authoritarians and nonauthoritarians were paired with other nonauthoritarians. In general the authoritarians tended to make accurate judgments about the *F*-scale scores of the other authoritarians with whom they had been paired, whereas the nonauthoritarians tended to make inaccurate judgments about the other nonauthoritarians. This result directly contradicts the assertion that nonauthoritarians are more insightful in interpersonal perception than authoritarians. Most revealing was the comparison that Scodel and Freedman provide of the two sets of results. In both studies, the *F*-scale estimates were highly related to the judge's own *F*-scale score but un-

related to the *F*-scale score of the partner with whom the judge was paired.¹ Authoritarians tended to estimate high *F*-scale scores for their partners, no matter what their actual score. Nonauthoritarians, however, tended to make more varied estimates of their partners, sometimes estimating high scores, sometimes low scores, but most often scores in the middle range.

Apparently, *F*-scale scores judged on the basis of a brief interaction reflect the assumptions of the judge, not the objective characteristics of the *S* judged. The findings of the Scodel-Mussen and Scodel-Freedman studies may arise from a difference in the beliefs held by authoritarians and nonauthoritarians about the *F*-scale responses of an "average" college student.

Gage and Cronbach, discussing conceptual and methodological problems in interpersonal perception, point out:

If we are asking a Judge to predict the response of an individual to a personality inventory, we...

¹ Since preparing this article, the writer's attention has been called to a study by Crockett and Meidinger (2) in which all combinations of subjects were used: authoritarian-authoritarian, nonauthoritarian-authoritarian, and nonauthoritarian-nonauthoritarian. Their results add further support to the assertion that the judge's estimate is not related to the actual *F*-scale score of his partner.

probably . . . want to measure "ability to predict how this individual deviates from the typical behavior of the particular group he belongs to." If accuracy is scored directly by comparing the prediction to the response of the individual, we are not distinguishing between two components which contribute to the Judge's success: his knowledge of the response that any individual in the subgroup is likely to give, and his knowledge of the way in which *this* individual deviates from the norm.

It is apparently desirable, when studying ability to predict at any one level, to obtain at least two scores: (a) ability to predict the typical behavior in the next-larger class to which Other belongs, and (b) ability to predict how Other deviates from the norm for this class (3, p. 417).

In neither of the studies under examination were Ss asked to predict the F-scale responses typical in their college student group. This note describes a study undertaken without knowledge of Scodel and Freedman's results, in which Ss with known F-scale scores marked the F scale as they believed a typical member of their college group would respond to it.

A group of 104 undergraduate college students (16 males, 88 females) were given a 28-item version of the F scale² twice: in the first administration, customary instructions were used, but in the second administration, Ss were asked to give the responses which they believed a typical student at their college would give. The second administration immediately followed the first. No S had access to his original F-scale responses during the second administration.

Table 1 summarizes the results for the total group of Ss, as well as the highest 20 per cent (authoritarians) and the lowest 20 per cent (non-authoritarians). The authoritarians had F-scale scores between 104 and 136, and the nonauthoritarians had F-scale scores between 36 and 61. Analysis of these data suggests the following conclusions:

1. For the total group, the difference between the means of the first and second administrations is highly significant ($t = 3.54, p < .001$). There is apparently a general tendency for these Ss to estimate higher F-scale scores for the typical college student than they themselves possess.

2. The mean estimate of the nonauthoritarians is approximately 38 points higher than their own mean, whereas the mean estimate of the authoritarians is approximately 9 points higher than their own mean. This difference is statistically highly significant ($t = 3.73, p < .001$). Apparently, the general tendency of the Ss to estimate F-scale

² The following F-scale items were used: Form 78, Items 6, 66, 67; Form 60, Items 22, 41, 50; Form 45 and 40, Items 1, 2, 6, 9, 12, 13, 18, 19, 21, 23, 26, 27, 29, 31, 33, 34, 37, 38, 41, 42, 43, 44 (1, pp. 226-227, 248-250, 255-257).

TABLE 1
F-SCALE SCORES OBTAINED UNDER TWO CONDITIONS OF INSTRUCTION

Instructions	Total Group (<i>N</i> = 104)		Non authori- tarians (<i>N</i> = 21)		Authori- tarians (<i>N</i> = 21)	
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>
Own attitude	83.1	23.4	51.3	8.0	116.0	9.3
Typical student	104.8	27.2	89.1	27.8	125.2	17.5

scores higher than their own is more marked among the nonauthoritarians than the authoritarians.

3. The mean estimate of the nonauthoritarians is only approximately 6 points higher than the actual total group average, whereas the mean estimate of the authoritarians is approximately 42 points higher than the actual group average. This difference is statistically highly significant ($t = 5.03, p < .001$). The nonauthoritarians are apparently more accurate than the authoritarians in estimating the true group average. Although this may reflect greater sensitivity to group opinion on the part of the nonauthoritarians, it seems more probable that it is a product of the general tendency of all Ss to estimate scores higher than their own, which happens, in this case, to make the estimates of the nonauthoritarians more accurate.

4. The estimates of the nonauthoritarians are more variable than those of the authoritarians ($F = 2.51; .01 < p < .05$). Apparently the authoritarians share a more uniform conception of the typical student's responses than do the non-authoritarians.

5. The correlation between the first and second administrations is highly significant ($r = .51, p < .001$), indicating the relative stability with which Ss maintain their rank position in the two distributions. Apparently, responding as a typical college student, low-scoring Ss tend to estimate low scores, and high-scoring Ss tend to estimate high scores.

Although a direct comparison is precluded, the results summarized in Table 1, based on Ss' estimates of the F-scale responses of a typical college student, are consistent with those of the Scodel-Mussen and Scodel-Freedman experiments. The experimental evidence demonstrates that judges who estimate the F-scale responses of others, after brief interpersonal contact, are unable to differentiate accurately among the others. The estimates depend more upon the F-scale score of the judge than that of the other. As in the cited experiments, the evidence of this study indicates that an S's estimate of the F-scale responses of a typical student depends largely upon the S's own

F-scale score. The combined data suggest two plausible hypotheses: (a) a judge who estimates the F-scale responses of an other, after brief interpersonal contact, does so by assuming the other is a typical member of the group to which he belongs, and (b) a judge's estimates of the F-scale responses of a typical member of a group depend more upon the judge's own F-scale responses than those characteristic of the group.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

2. CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and interpersonal perception. *J. abnorm. soc. Psychol.*, 53, 378-380.
3. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, 62, 411-422.
4. SCODEL, A., & FREEDMAN, MARIA L. Additional observations on the social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1956, 52, 92-95.
5. SCODEL, A., & MUSSEN, P. Social perceptions of authoritarians and nonauthoritarians. *J. abnorm. soc. Psychol.*, 1953, 48, 181-184.

Received February 27, 1956.

A COMPARISON OF THE DRQ WITH RATINGS OF EMOTION¹

FRANK AULD, JR. AND GEORGE F. MAHL

Yale University

THE Discomfort-Relief Quotient, or DRQ, was invented by Dollard and Mowrer (3, 4) as a method of measuring tension in written documents. To obtain the DRQ, a scorer classifies each unit of the written material as expressive of discomfort, relief, or neutral feeling, and then computes the ratio of discomfort units to the total of discomfort and relief units. Transcribed interviews, as well as communications originally made in written form, can be scored by this method.

The question naturally arises whether the quotient so obtained accurately measures the tension felt by the communicator. Meadow and his co-workers (8), believing that the DRQ might not accurately reflect psychological tension, compared the DRQ scores of 35 schizophrenic patients with a psychiatrist's ratings of the tension felt by these patients. Meadow *et al.* found a negligible and insignificant correlation between the DRQ scores and the ratings of tension. However, this result does not prove that the DRQ is invalid, because: (a) The ratings of tension were made in a different situation from that in which material for the DRQ scores was gathered; the lack of correlation may only show that patients behave differently in different situations. (b) Meadow *et al.* computed the DRQ erroneously, including all units in the denominator. (c) Since the subjects were schizophrenic, the results may only show that the DRQ

is invalid for psychotics, who are known (2) to verbalize their feelings less accurately than normals and neurotics.

The aim of the present study is to appraise the validity of the DRQ as a measure of tension, using as subjects a group consisting mostly of neurotics, and obtaining the ratings of tension from the same situation that provides the material for the DRQ scores.

METHOD

Subjects. The Ss are 39 persons who either applied for treatment in the Psychiatric Outpatient Clinic of the New Haven Dispensary or entered the Yale Psychiatric Institute, a private psychiatric hospital, as patients. Of the 39 patients, 31 are outpatients and 8 are inpatients. Twenty-two of the Ss were diagnosed as psychoneurotic, 10 as suffering from character disorders, and 7 as psychotic.

Procedure. Three interviewers—a psychologist and two psychiatric residents—took turns interviewing the Ss, one of the team doing the interviewing while the others monitored and observed through a one-way mirror.² For the outpatients these were initial interviews, conducted in order to permit a diagnostic formulation; for the inpatients these are special research interviews. Immediately after each interview all members of the team made independent ratings of the patient's behavior, on a five-point scale, on each of three variables: anxiety, hostility, and dependence. All interviews were recorded on magnetic tape, were later carefully transcribed, and were divided into sentence units according to rules developed by Auld, Dollard, and White (1). The two authors and a secretary scored the interviews according to the DRQ method, one author (G.F.M.) scoring the entire series,

¹ The authors wish to thank the National Institute of Mental Health, U. S. Public Health Service, for financial support of this investigation, through grant M-648, "Development of Quantitative Methods for the Study of Psychotherapy" (Professor John Dollard, principal investigator), and grant M-1052, "The Patient's Language as Expressive Behavior" (George F. Mahl, principal investigator). The authors are also grateful for financial aid from the Foundations Fund for Research in Psychiatry.

² These interviews and ratings were obtained for other purposes by Drs. George Andrews, George F. Mahl, and Louis Micheels. The facilities for the recordings and observation of these interviews have been described elsewhere (7). Additional information about the interviews is presented in a previous paper (5).

TABLE 1
RELIABILITY OF RATINGS OF EMOTION

Variable	Interjudge Reliability Coefficients		
	A with B	A with C	B with C
Anxiety	.68****	.22	.42***
Hostility	.61****	.46***	.47***
Dependence	.47***	.37**	.28*

- * Significant at .10 level.
 ** Significant at .02 level.
 *** Significant at .01 level.
 **** Significant at .001 level.

TABLE 2
CORRELATIONS OF COMPOSITE RATINGS WITH DRQ

Variable	Correlation Coefficients		
	Men (N = 13)	Women (N = 26)	Total Sample
Anxiety	-.11	.33*	.27*
Hostility	.02	.15	.12
Dependence	.24	.37*	.31*
Anxiety + Hostility + Dependence	.15	.41**	.35**

- * Significant at .10 level.
 ** Significant at .05 level.

TABLE 3
CORRELATIONS OF COMPOSITE RATINGS WITH DRQS
OF TWO SCORERS

Variable	Correlation Coefficient*	
	DRQ of F.A.	DRQ of G.F.M.
Anxiety	.08	.17
Hostility	.10	.04
Dependence	.24	.36
Anxiety + Hostility + De- pendence	.21	.28

- * With a sample of only 21 cases, none of the reported coefficients is statistically significant at the .10 level.

the other (F.A.) scoring 21 interviews, and the secretary scoring 10 interviews.

It was planned to have the secretary score all the interviews and to use her DRQ scores in the analysis of results, provided that she should score reliably. When it turned out that her scores had a correlation of only .51 with G.F.M.'s and .53 with F.A.'s, the authors decided to use G.F.M.'s scores instead.

RESULTS

Reliability. The agreement among judges for the ratings of anxiety, hostility, and dependence is reported in Table 1. To obtain the most reliable single measure of anxiety and hostility, we added A's and B's ratings, ignoring C's. To obtain the most reliable measure of dependence, we took the sum of all three ratings. As estimated by the Spearman-Brown formula the reliability of the

anxiety composite is .81, of the hostility composite, .76, and of the dependence composite, .62.

Reliability of the division of interviews into sentences was established by comparing the secretary's unitizing with that of the authors. There was agreement on placement of more than 95 per cent of the unit markings.

The correlation between the authors' DRQ scores of the 21 hours scored by both was .90, which is statistically significant at the .001 level.

Correlation of ratings with DRQ. Table 2 shows the correlations of the composite ratings with the DRQ, for men and for women separately and for the total sample. Taken all in all, there is a slight positive relationship between the ratings on these variables and the DRQ.

The correlations between the ratings and the DRQ are consistently higher for the women than for the men. The probability is about .22, as figured by Stouffer's method of combining probabilities (9), that three such differences between correlations for men and for women as we have observed would occur simultaneously by chance. A study by Mahl (6) of the correlation between disturbances in speech and the anxiety ratings has revealed a similar and highly significant sex difference. The meaning of these sex differences is obscure at the present time.

Possibility of contamination. Since one of the raters was also the person whose DRQ scores were reported above, it is possible that any correlation between the ratings and the DRQ scores is spurious. To exclude this possibility we separately analyzed the 21 interviews which the other author had scored. The results of this analysis are presented in Table 3. No conclusion of this study would be changed if F.A.'s scores rather than G.F.M.'s were used.

DISCUSSION

It does not seem likely that the small size of the correlations between the global ratings and the DRQ can be accounted for by poor reliability of the measures. When the correlations are corrected for attenuation they are still of unimpressive size. To the degree that the global ratings are themselves valid, therefore, the failure of the DRQ scores to correlate with them more highly should discourage the use of the DRQ as a measure of individual differences in anxiety, hostility, or dependence.

It must be noted that this conclusion applies, strictly, only to use of the DRQ as a measure of *individual differences* in tension. The findings do not tell whether a change in the DRQ within an interview series of a single patient would validly measure changes in the tension felt by the patient.

The reader's attention is directed to the high reliability of the DRQ scores. Whereas interjudge

reliabilities ranged from .22 to .68 for the global ratings of emotion, the interjudge reliability of the two authors' DRQ scores was .90. The superior reliability of the DRQ procedure suggests that one might profitably seek to increase the reliability of measures of anxiety, hostility, dependence, and the like by adopting scoring procedures patterned on the DRQ. Perhaps such procedures, based on content-analysis methods and including behavior items and vocal as well as verbal cues, could measure essentially the same variables as the global judgments but do so with greater precision.

SUMMARY

In order to appraise the validity of the DRQ as a measure of tension, the authors compared DRQ scores of 39 psychiatric interviews with global ratings of the same interviews on anxiety, hostility, and dependence. There were small positive correlations between the global ratings and the DRQ, these correlations tending to be higher for the women than for the men. If the global ratings can be accepted as adequate measures of tension, the DRQ does not measure individual differences in tension very well. However, one must keep in mind that this study focuses on measurement of individual differences in tension and is not immediately pertinent to use of the DRQ as a measure of changes in tension within a series of interviews with a single patient.

REFERENCES

1. AULD, F., DOLLARD, J., & WHITE, ALICE M. A method of dividing interviews into independent-clause units. Paper read at East. Psychol. Assoc., Philadelphia, April 16, 1955.
2. DOLLARD, J. The psychotic person seen culturally. *Amer. J. Sociol.*, 1934, **39**, 637-648.
3. DOLLARD, J., & MOWRER, O. H. Progress report on casework research. New York: Community Service Society, 1943.
4. DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3-32.
5. LAMB, R., & MAHL, G. F. Manifest reactions of patients and interviewers to the use of sound recording in the psychiatric interview. *Amer. J. Psychiat.*, 1956, **112**, 731-737.
6. MAHL, G. F. Disturbances in the patient's speech as a function of anxiety. Paper read at East. Psychol. Assoc., Atlantic City, March 23, 1956.
7. MAHL, G. F., DOLLARD, J., & REDLICH, F. C. Facilities for sound recording and the observation of interviews. *Science*, 1954, **120**, 235-239.
8. MEADOW, A., GREENBLATT, M., LEVINE, J., & SOLOMON, H. C. The Discomfort-Relief Quotient as a measure of tension and adjustment. *J. abnorm. soc. Psychol.*, 1952, **47**, 658-661.
9. MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289-334.

Received June 29, 1956. Prior publication.

A SIMPLE DEVICE FOR OBTAINING CERTAIN VERBAL ACTIVITY MEASURES DURING INTERVIEWS¹

STANISLAV V. KASL AND GEORGE F. MAHL

Yale University

DURING the past decade and a half there have been many objective studies of interview transactions. The vast majority of these have consisted of verbal content analyses, i.e., the reliable description in quantitative terms of "what" the participants have said. The "how" (6), or the "vocal" (7), aspects of the communication have received limited attention in spite of the common knowledge that they are just as important as the verbal content, and often more so.

One particular class of the "how" of verbal activity includes those features of speech that are exclusively behavioral (i. e., without conventional semantic function), such as the amount of talking by the participants in standard time intervals, the rate of talking when actual speaking occurs, the amount of silence, interruptions, etc. Any

extensive study of such verbal activity, however, requires some instrumentation beyond a stop watch. Chapple (1, 2) devised his Interaction Chronograph and Verzeano and Finesinger (9) their Automatic Speech Analyzer for such purposes. Matarazzo *et al.* (5) give a comprehensive picture of the functioning of the Chapple Interaction Chronograph, the data obtained with it, and of studies based on its use. Applications of the Verzeano-Finesinger device are found in papers by Verzeano (8), and Lorenz and Cobb (3).

During the course of investigating the expressive aspects of the patient's language in psychotherapy, it became necessary to determine the duration of time spent talking by the participants, and the amount of time the patient was silent during successive two-minute intervals of individual interviews as well as during larger and variable time intervals. An instrument for these relatively restricted purposes does not have to be as elaborate as the Interaction Chronograph or Speech Analyzer which yield additional data, and yet it requires

¹ This study is a fragment of a research program supported by USPHS Grant M-1052, The Patient's Language as Expressive Behavior (George F. Mahl, principal investigator), and by the Foundations' Fund for Research in Psychiatry.

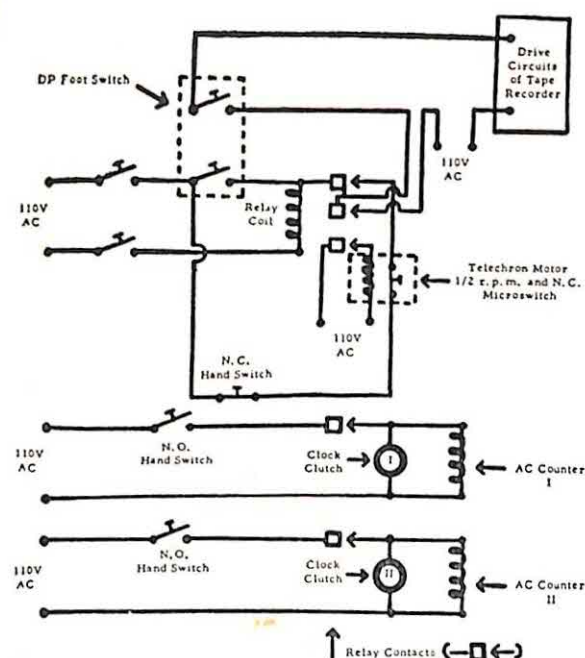


FIG. 1. CIRCUIT DIAGRAM OF TIMING INSTRUMENT

certain features not found in them. Consequently, one of the writers (G.F.M.) devised the simple instrument that is described in this paper. This is essentially a development of Chapple's original idea for a special purpose. The instrument is inexpensive, has been of practical value, and its use yields reliable data. Others may also find it to be helpful.

Figure 1 contains the circuit diagram. The instrument is for use with tape recordings of interviews and playback equipment that is started and stopped with electronic rather than mechanical controls. How the instrument is connected with the electronic controls of the driving mechanism of the playback equipment may vary in detail with the particular machine that is used. Since there is nothing special about the parts used in the apparatus, no detailed specifications are presented.

An observer operates the apparatus while listening to the recording through earphones. In obtaining the successive two-minute measures, the instrument functions as follows. The momentary closure of the observer's foot-switch starts the tape recorder, a telechron motor geared for $\frac{1}{2}$ r.p.m., and closes the relay switches in series with the clutches of two accumulating clocks (Spring-field timers) and the coils of two counters. The operator may then release his foot and the various elements will still be activated since the relay is wired as a holding relay. Now the observer operates the two hand-switches (microswitches equipped with a metal button on the switch arm for sensitive finger operation) that are also in series with the clock clutches and the counter coils. When he now presses them each time the patient

or the therapist talks, the appropriately assigned clock will run for as long as the key is depressed and the counter will make one count per closure and opening of the switch (this is called an "act"). The "Normally Closed Microswitch" is opened when the telechron motor completes one revolution in two minutes. This instantly releases the relay, opening all the relay contacts so as to stop the tape recorder and inactivate the clocks and counters. The clocks and counters are then read and zeroed. Starting again with a momentary closure of the foot-switch, the same sequence is now repeated for the next two minute. By using this apparatus the raw data for a 50-60 minute interview, consisting of the number of seconds of talking and the number of "acts" for each participant in each two-minute interval, can be obtained in about one and a half hours by an experienced operator. Practice runs on 10 to 15 interviews give an inexperienced person sufficient training.

If one wishes to use a fixed time segment other than of two minutes duration, all that is required is to use a telechron motor that is geared to make 1 r.p.m. in the preferred time interval.

For some purposes it is desirable to obtain the activity measures for variable time segments, e.g. during a period when the patient is talking about various specific topics, or when he is engaged in a particular defense, etc. This is done with the present device by short circuiting the microswitch

TABLE 1
TEST-RETEST AND INTEROBSERVER CORRELATIONS OF
TWO-MINUTE ACTIVITY MEASURES IN 5 UNSELECTED
INTERVIEWS*

Activity Measure	Observers Compared	Product Moment Correlations	
		\bar{X}	Range
Seconds of Patient Talk.	A-A	.98	.957-.992
	B-B	.97	.923-.986
	A-B	.97	.940-.986
Seconds of Therapist Talk.	A-A	.99	.972-.997
	B-B	.98	.965-.986
	A-B	.99	.972-.996
N Patient "Acts"	A-A	.88	.455-.954
	B-B	.78	.602-.879
	A-B	.76	.566-.910
N Therapist "Acts"	A-A	.94	.928-.954
	B-B	.92	.796-.964
	A-B	.94	.882-.971
Silence Quotient	A-A	.97	.943-.990
	B-B	.91	.817-.949
	A-B	.94	.896-.973

* These correlations reflect the observations of Jon Eklund and Stanislav Kasl.

Observer A's retests made 2-4 weeks after initial test.

Observer B's retests made 6 months after initial runs of these 5 and 34 other interviews with no intervening practice on other material.

operated by the telechron motor. The tape is set to start with the onset of the particular phase. The operator then starts with the momentary closure of the foot-switch. But now, with the telechron microswitch shorted, the apparatus will run until the "normally closed hand-switch" in Fig. 1 is opened. We have used a second observer to monitor the tape along with the operator and to open this hand-switch at the end of the particular phase. The operator is thus left free to attend only to the ongoing verbal behavior in the tape, so that he is not distracted by having the additional task of "listening for" the end of the phase.

Three people have learned independently to operate the apparatus, and in each case they developed similar habits to maximize their efficiency and accuracy. It was felt necessary, for example, to concentrate on the speech, yet pay no attention to its actual content or meaning. Eyes were fixed on some neutral object in the room or on the hand of the Springfield timer, perhaps to restrict the range of attention. The resulting set seemed to facilitate the observers' responding to the speech as an irregular, intermittent auditory stimulus, rather than to such linguistic units as phrases and sentences.

We have used the activity measures in two ways. The number of seconds of talking has been used with word counts to obtain "rates of talking when talking." Also a Patient Silence Quotient has been determined for successive two-minute intervals of therapeutic interviews, and in various motivational phases of the interview. The Patient Silence Quotient² is defined as:

Patient Silence Quotient =

$$\frac{N \text{ seconds of silence}}{N \text{ seconds available to patient to talk}}$$

$a = N$ seconds in the segment (120" e.g. in 2')

$b = N$ seconds talk by therapist

$c = N$ seconds talk by patient

N seconds of silence = $a - (b + c)$

N seconds available to patient = $a - b$.

The count of "acts" has not yet been used. But it is possible that they will help provide a measure of the speaker's fluency, independent of the Silence Quotient. Table 1 contains test-retest and inter-observer correlations for the various activity measures in two-minute intervals of 5 unselected interviews. The N s ranged from 25 to 30 two-minute intervals.

² It can be seen from the definition of the Silence Quotient that sometimes brief hesitations, "naturally" attributable to the therapist, that is, in which the therapist gives every indication of continuing his talk but does not do so for a brief moment, are scored as "time available to patient." The amount of time that is thus erroneously treated is regarded as negligible.

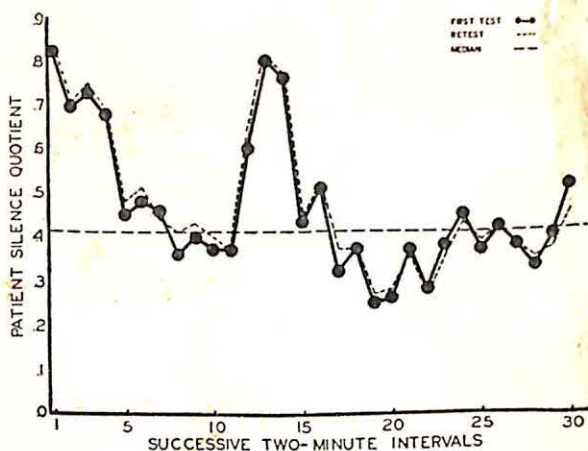


FIG. 2. TEST-RETEST PATIENT SILENCE QUOTIENTS IN SUCCESSIVE TWO-MINUTE INTERVALS OF A THERAPEUTIC INTERVIEW. MEDIAN IS OF FIRST SET OF MEASURES

Figure 2 is a graph of the Silence Quotient in successive two-minute intervals of a therapeutic interview. This graph illustrates the test-retest reliability, as well as the clear-cut depiction of behavioral change during the course of an interview that is possible by means of the measurement of "how" the patient talks. A previous paper (4) has dealt with some of the broader implications of such data.

REFERENCES

1. CHAPPLE, E. D. Measuring human relations: an introduction to the study of interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3-147.
2. CHAPPLE, E. D. The interaction chronograph: its evolution and present application. *Personnel*, 1949, **25**, 295-307.
3. LORENZ, M., & COBB, S. Language behavior in manic patients. *Arch. Neurol. Psychiat.*, 1952, **69**, 763-770.
4. MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1-15.
5. MATARAZZO, J. D., SASLOW, G., & MATARAZZO, R. G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347-367.
6. REICH, W. On character analysis. In R. Fliess (Ed.), *The psychoanalytic reader*. New York: International Universities Press, 1948.
7. SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954.
8. VERZEANO, M. Time-patterns of speech in normal subjects. *J. Speech Hearing Disorders*, 1950, **15**, 197-201.
9. VERZEANO, M., & FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 1949, **110**, 45-46.

Received July 24, 1956. Prior Publication.

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, New York University

Associate Editor: E. J. SHOBN, JR., Teachers College, Columbia University

Managing Editor: LORRAINE BOUTHILET

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
The RAND Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

LEON FESTINGER
Stanford University

DONALD W. FISKE
University of Chicago

ROBERT R. HOLT
New York University

HERBERT H. HYMAN
Columbia University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
Center for Advanced Study
in the Behavioral Sciences

GARDNER LINDZEY
Syracuse University

DAVID C. MCCLELLAND
Harvard University

QUINN MCNEMAR
Stanford University

LEO POSTMAN
University of California

NEVITT SANFORD
Vassar College

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

VOLUME 53
1956



PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.

AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Copyright 1956 by the American Psychological Association, Inc.

CONTENTS OF VOLUME 53

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

	PAGE
ALLPORT, F. H. See TANNENBAUM, A. S.	
ANESHANSEL, J. See JONES, E. E.	
ARNHOFF, F. N. Ethnocentrism and Stimulus Generalization.....	138
ATKINSON, J. W., & REITMAN, W. R. Performance as a Function of Motive Strength and Expectancy of Goal-Attainment.....	361
ATKINSON, J. W., & WALKER, E. L. The Affiliation Motive and Perceptual Sensitivity to Faces.....	38
AULD, F., JR., & MAHL, G. F. A Comparison of the DRQ With Ratings of Emotion.....	386
BARRY, J. R. A Further Study of the McKinney Reporting Test Scores.....	258
BASS, B. M. Development and Evaluation of a Scale for Measuring Social Acquiescence.....	296
BECKER, W. C. A Genetic Approach to the Interpretation and Evaluation of the Process-Reactive Distinction in Schizophrenia.....	229
BEIER, H., & HANFMANN, E. Emotional Attitudes of Former Soviet Citizens, as Studied by the Technique of Projective Questions.....	143
BELLAMY, R. Q. See BLAKE, R. R.	
BERKOWITZ, H. See BLAKE, R. R.	
BERKOWITZ, L., & LEVY, B. I. Pride in Group Performance and Group-Task Motivation.....	300
BERNSTEIN, L. A Reply to Scott.....	141
BEVAN, W. See SECORD, P. F.	
BIERI, J., & BLACKER, E. The Generality of Cognitive Complexity in the Perception of People and Inkblots.....	112
BINDRA, D. Psychotherapy and the Recovery From Neurosis.....	251
BLACKER, E. See BIERI, J.	
BLAKE, R. R., BERKOWITZ, H., BELLAMY, R. Q., & MOUTON, J. S. Volunteering as an Avoidance Act..	154
BROOKS, L. O. See MALTZMAN, I.	
BROWN, J. E. See COWDEN, R. C.	
BROWN, R. W. See HILDUM, D. C.	
BYRNE, D. The Relationship Between Humor and the Expression of Hostility.....	84
CAMPBELL, D. T. Enhancement of Contrast as Composite Habit.....	350
CARTER, L. See HAYTHORN, W.	
CASTANEDA, A. Reaction Time and Response Amplitude as a Function of Anxiety and Stimulus Intensity.....	225
CHAMBERS, G. S., & ZABARENKO, R. N. Effects of Glutamic Acid and Social Stimulation in Mental Deficiency.....	315
CHAPMAN, L. J. Distractibility in the Conceptual Performance of Schizophrenics.....	286
CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope of Success and Fear of Failure as Aspects of Need for Achievement.....	182
COHEN, D. See NIEBUHR, H., JR.	
COUCH, A. See HAYTHORN, W.	
COWDEN, R. C., & BROWN, J. E. The Use of a Physical Symptom as a Defense Against Psychosis...	133
CROCKETT, W. H., & MEIDINGER, T. Authoritarianism and Interpersonal Perception.....	378
CROMWELL, R. L. Factors in the Serial Recall of Names of Acquaintances.....	63
DAVIDS, A., JOELSON, M., & MCARTHUR, C. Rorschach and TAT Indices of Homosexuality in Overt Homosexuals, Neurotics, and Normal Males.....	161
DESOTO, C., & LEIBOWITZ, H. Perceptual Organization and Intelligence: A Further Study.....	334
DITTES, J. E., & KELLEY, H. H. Effects of Different Conditions of Acceptance Upon Conformity to Group Norms.....	100
DORIS, J., & FIERMAN, E. Humor and Anxiety.....	59
DOWNEY, J. F. See KOGAN, N.	
ERIKSEN, C. W., & KUETHE, J. L. Avoidance Conditioning of Verbal Behavior Without Awareness: A Paradigm of Repression.....	203
EYSENCK, H. J. Reminiscence, Drive, and Personality Theory.....	328
FIERMAN, E. See DORIS, J.	
FIRTH, R. Rumor in a Primitive Society.....	122
FRENCH, E. G. Motivation as a Variable in Work-Partner Selection.....	96
RUCHTER, B. See ROKEACH, M.	
ILBERT, D. C., & LEVINSON, D. J. Ideology, Personality, and Institutional Policy in the Mental Hospital.....	263
MILCHRIST, J. C. See HANSCHKE, J.	
MAEFNER, D. See HAYTHORN, W.	

	PAGE
HANFMANN, E. See BEIER, H.	
HANSCH, J., & GILCHRIST, J. C. Three Determinants of the Level of Aspiration.....	136
HAYTHORN, W., HAEFNER, D., LANGHAM, P., COUCH, A., & CARTER, L. The Effects of Varying Combinations of Authoritarian and Equalitarian Leaders and Followers.....	210
HILDUM, D. C., & BROWN, R. W. Verbal Reinforcement and Interviewer Bias.....	108
JACKSON, D. N. A Short Form of Witkin's Embedded-Figures Test.....	254
JAHODA, G. Assessment of Abstract Behavior in a Non-Western Culture.....	237
JOELSON, M. See DAVIDS, A.	
JONES, A. Distributions of Traits in Current Q-Sort Methodology.....	90
JONES, A. Sexual Symbolism and the Variables of Sex and Personality Integration.....	187
JONES, E. E., & ANESHANSEL, J. The Learning and Utilization of Contravalant Material.....	27
JONES, L. V., & MORRIS, C. Relations of Temperament to the Choice of Values.....	346
KAGAN, J. The Child's Perception of the Parent.....	257
KAROLCHUK, P. A., & WORELL, L. Achievement Motivation and Learning.....	255
KASL, S. V., & MAHL, G. F. A Simple Device for Obtaining Certain Verbal Activity Measures During Interviews.....	388
KATZ, B. See SECORD, P. F.	
KEEHN, J. D. See PROTHRO, E. T.	
KELLEY, H. H. See DITTES, J. E.	
KLUGMAN, S. F. Retention of Affectively Toned Verbal Material by Normals and Neurotics.....	321
KOGAN, N. Authoritarianism and Repression.....	34
KOGAN, N., & DOWNEY, J. F. Scaling Norm Conflicts in the Area of Prejudice and Discrimination..	292
KUETHE, J. L. See ERIKSEN, C. W.	
LAMBERT, W. E., & TAGUCHI, Y. Ethnic Cleavage Among Young Children.....	380
LANGHAM, P. See HAYTHORN, W.	
LANZETTA, J. T., & ROBY, T. B. Effects of Work-Group Structure and Certain Task Variables on Group Performance.....	307
LEIBOWITZ, H. See DESOTO, C.	
LEVIN, H. See MACCOBY, E. E.	
LEVINSON, D. J. See GILBERT, D. C.	
LEVY, B. I. See BERKOWITZ, L.	
LEVY, L. H. Personal Constructs and Predictive Behavior.....	54
MCCARTHY, C. See DAVIDS, A.	
MACCOBY, E. E., LEVIN, H., & SELYA, B. M. The Effects of Emotional Arousal on the Retention of Film Content: A Failure to Replicate.....	373
MAHER, H. See FREEMAN, J. T.	
MAHL, G. F. Disturbances and Silences in the Patient's Speech in Psychotherapy.....	1
MAHL, G. F. See AULD, F., JR.	
MAHL, G. F. See KASL, S. V.	
MALTZMAN, I., MORRISETT, L., JR., & BROOKS, L. O. An Investigation of Phonetic Symbolism.....	249
MEIDINGER, T. See CROCKETT, W. H.	
MILLER, R. E. See MURPHY, J. V.	
MONACHESI, E. D. See OLMSTED, D. W.	
MONKMAN, J. A. See STANLEY, W. C.	
MORRIS, C. See JONES, L. V.	
MORRISETT, L., JR. See MALTZMAN, I.	
MOUTON, J. S. See BLAKE, R. R.	
MURPHY, J. V., & MILLER, R. E. The Manipulation of Dominance in Monkeys With Conditioned Fear.....	244
NASS, M. L. The Effects of Three Variables on Children's Concepts of Physical Causality.....	191
NIEBUHR, H., JR., & COHEN, D. The Effect of Psychopathology on Visual Discrimination.....	173
OLMSTED, D. W., & MONACHESI, E. D. A Validity Check on MMPI Scales of Responsibility and Dominance.....	140
OPLER, M. K. See SINGER, J. L.	
PAUL, I. H. Impressions of Personality, Authoritarianism, and the <i>Fait-Accompli</i> Effect.....	338
PROTHRO, E. T., & KEEHN, J. D. The Structure of Social Attitudes in Lebanon.....	157
RABINOWITZ, W. A Note on the Social Perceptions of Authoritarians and Nonauthoritarians.....	384
RADCLIFFE, K. B., JR. See WITTEICH, W. J.	
REITMAN, W. R. See ATKINSON, J. W.	
RICCIUTI, H. N. See CLARK, R. A.	
ROBY, T. B. See LANZETTA, J. T.	
ROCKETT, F. C. Speed of Form Recognition as a Function of Stimulus Factors and Test Anxiety..	197
ROKEACH, M., & FRUCHTER, B. A Factorial Study of Dogmatism and Related Concepts.....	356

	PAGE
ROSENBAUM, G. Stimulus Generalization as a Function of Clinical Anxiety.....	281
ROSENBAUM, M. E. The Effect of Stimulus and Background Factors on the Volunteering Response.....	118
ROSENBERG, M. J. Cognitive Structure and Attitudinal Affect.....	367
SECORD, P. F., BEVAN, W., & KATZ, B. The Negro Stereotype and Perceptual Accentuation.....	78
SELYA, B. M. See MACCOBY, E. E.	
SINGER, J. L., & OPLER, M. K. Contrasting Patterns of Fantasy and Motility in Irish and Italian Schizophrenics.....	42
SINGER, J. L., & SPOHN, H. E. The Response of Schizophrenic Patients to a Televised World Series Game: A Study in Social Isolation.....	375
SMOCK, C. D. Replication and Comments: "An Experimental Reunion of Psychoanalytic Theory With Perceptual Vigilance and Defense.....	68
SPOHN, H. E. See SINGER, J. L.	
STANLEY, W. C., & MONKMAN, J. A. A Test for Specific and General Behavioral Effects of Infantile Stimulation With Shock in the Mouse.....	19
TAGUCHI, Y. See LAMBERT, W. E.	
TANNENBAUM, A. S., & ALLPORT, F. H. Personality Structure and Group Structure: An Interpretative Study of Their Relationship Through an Event-Structured Hypothesis.....	272
TEEVAN, R. See CLARK, R. A.	
TEMERLIN, M. K. One Determinant of the Capacity to Free-Associate in Psychotherapy.....	16
TIZARD, J. See VENABLES, P. H.	
VENABLES, P. H., & TIZARD, J. Paradoxical Effects in the Reaction Time of Schizophrenics.....	220
VENABLES, P. H., & TIZARD, J. Performance of Functional Psychotics on a Repetitive Task.....	23
WALKER, E. L. See ATKINSON, J. W.	
WILLIS, R. H. Political and Child-Rearing Attitudes in Sweden.....	74
WITTEICH, W. J., & RADCLIFFE, K. B., JR. Differences in the Perception of an Authority Figure and a Nonauthority Figure by Navy Recruits.....	383
WORELL, L. The Effect of Goal Value Upon Expectancy.....	48
WORELL, L. See KAROLCHUCK, P. A.	
YATES, A. J. The Rotation of Drawings by Brain-Damaged Patients.....	178
ZABARENKO, R. N. See CHAMBERS, G. S.	





